University of Alberta

Ability and Achievement Variables of Average, Low Average, and Borderline Students

by

Timothy Robert Claypool ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the

requirements for the degree of Doctor of Philosophy

in

School Psychology

Department of Educational Psychology

Edmonton, Alberta

Spring 2005

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

0-494-08220-8

Your file Votre référence
ISBN:
Our file Notre reference
ISBN:

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Students are differentiated semantically and psychometrically when their levels of intelligence are determined through tests, such as the Wechsler Intelligence Scales for Children – Third Edition (WISC-III). The two-fold purpose of this study was to combine semantic and psychometric information in a manner that is not typically found in the research literature, with the intent of contributing to the evolution of clinical practice in school psychology. To this end, the IQ-based classifications of Average, Low Average, and Borderline were examined to determine empirical similarities and differences in levels of achievement in word reading and mathematics calculation. Additionally, psychometric manipulations of IQ were permitted through the use of the 'Symbol Search substitution' method and an eight subtest abbreviated IQ in the form of the General Ability Index. The four psychometric-based factor scores were also used as a means of comparison for the three IQ-based classifications.

The 196 cases that comprised the archival research sample completed psychoeducational assessment through the University of Alberta's Education Clinic. Graduate level students enrolled in the Educational Psychology course, EDPY 545, completed these assessments under the close supervision of the clinic's director, and his appointed supervisors. The subject pool consisted of 88 females and 108 males that were distributed among the three IQ classifications.

When a 3 X 2 between subjects factorial design was employed, gender did not prove to be a significant distinguishing variable for academic achievement levels among IQ-based classifications. This result was consistent for all three methods of

calculating IQ. Similarly, age groupings were not a significant variable when reading and mathematics levels were compared across the Average, Low Average and Borderline groupings.

In all cases, the Low Average and Borderline groups' achievement levels differed significantly from that of the Average group. The fact that reading and mathematics abilities were not differentiated when Low Average and Borderline groups were compared, calls into question the veracity of these labels.

This research supports the drafting of a Canadian version of a 'Rights without Labels' policy statement, which is presently only evident in an American format courtesy of the National Association of School Psychologists.

Acknowledgements

Completion of this research study would not be possible without the ongoing support and guidance provided by my thesis advisor, Dr. H. L. Janzen. His practical knowledge combined with a commitment to improving assessment practices through informed research was both an inspiration and a model for change. I would also like to thank my thesis committee for their suggestions and support.

It goes without saying that the efforts of the Education Clinic staff, the graduate students enrolled in EdPy 545 and their supervisors provided the data needed to complete this study. The professional standards maintained at all levels of this dynamic process not only adds to validity of the results but made the 'Clinic' a great place to work.

Special acknowledgement is given to all the individuals who completed psychoeducational assessments at the Education Clinic. Their anonymous contribution to ongoing research in this important area of School Psychology is valued.

Finally, I would like to recognize the contribution of my family and friends who were there when I needed them.

Table of Contents

Chapter 1

Introduction...1

Purpose of Study..1

Overview..3

Chapter 2

Literature Review...11

Politics of Assessment..11

Role of the School Psychologist..16

Intelligence, Cognition and IQ..20

Future Considerations...27

Summary..29

Chapter 3

The Study..32

Research Framework and Rationale..32

Research Hypotheses...33

Research Methodology..35

Testing Hypotheses One (a) and (b) and Three and Four..................35

Testing Hypotheses Two...38

Analysis of Data...38

Hypothesis One..39

Hypothesis Two..40

Hypothesis Three..40

Hypothesis Four..41

Wechsler Intelligence Scale for Children (WISC-III)..........................41

Chapter 4

Results..43

Overview..43

The Research Sample...43

Descriptive Results...51

Subtest Effect 52
Gender 53
Age and Grade Levels 54
Achievement Data Analysis 58
Reading versus Mathematics 59
Correlational Analysis 62
Gender and Correlation 63
Correlation and IQ Calculation Method 66
Inferential Statistics 68
Hypothesis One 68
Hypothesis Two 68
Hypothesis Three 69
Hypothesis Four 69
Testing Hypotheses One (a) and (b) 69
Hypothesis 1. (a) 70
Hypothesis 1. (b) 74
Testing Hypothesis Two 78
Testing Hypothesis Three 81
Testing Hypothesis Four 90
Summary 95
Correlation 95
Research Hypotheses 96
Hypothesis One 96
Hypothesis Two 97
Hypothesis Three 99
Hypothesis Four 99

Chapter 5

Discussion 102
Abstract 102
Delimitations 102
Rationale Revisited 104

Calculated Differences....105

Retention Effect – A Psychometric 'Fly in the Ointment'...107

Correlations of Interest...109

Speed of Processing and Intelligence...110

Schooling, Intelligence and Metacognitive Considerations...115

Theory, Research and Practice Connections...122

Conclusions...125

Closing Thoughts....128

References...132

Appendix...143

List of Tables

Table 4.1 Comparison of WISC-III FSIQ Classification and Educational Programming........49

Table 4.2 Comparison of WISC-III FSIQ Classification and Comorbid Attentional, Behavioral, and Emotional Conditions........49

Table 4.3 Summary of WISC-III Results for Sample and IQ-Based Classifications......51

Table 4.4 Scaled Score results of WISC-III Subtests for Sample and IQ-Based Classifications........53

Table 4.5 Achievement Test Administered According to IQ Classification........58

Table 4.6 Achievement Test Means and Standard Deviations for Sample (N = 195).....59

Table 4.7 Word Recognition and Mathematics Calculation Means and Standard Deviation for Full Sample (*N* = 196) and FSIQ Classifications........60

Table 4.8 Inter-correlations Between WISC-III Subtests........63

Table 4.9 Correlations Between IQ, Achievement and Gender........64

Table 4.10 Correlations Between IQ Classifications, Achievement and Gender..........65

Table 4.11 Correlations Between IQ Classifications, and Achievement........66

Table 4.12 Two-Way ANOVA (IQ Classification X Gender) for Word Reading Ability........70

Table 4.13 Two-Way ANOVA (IQ Classification X Gender) for Mathematics Calculation........73

Table 4.14 Two-Way ANOVA (IQ Classification X Age Group) for Word Reading Ability........75

Table 4.15 Two-Way ANOVA (IQ Classification X Age Group) for Mathematics Calculation........77

Table 4.16 Univariate *F* Tests for IQ Classification and WISC-III Factor Scores.........79

Table 4.17 Effect of Symbol Search Replacement FSIQ on IQ Classifications...........82

Table 4.18 Comparison of IQ Classifications between Standard FSIQ and FSIQss Methods........86

Table 4.19 Comparison of Reading Achievement Standard Score Means for Standard IQ and Symbol Search Substitution Classifications According to Gender........87

Table 4.20 Comparing Two-Way ANOVAs (IQ Classification X Gender) for Word Reading Utilizing Standard FSIQ and FSIQss.....88

Table 4.21 Comparison of Mathematics Achievement Standard Score Means and Standard Deviations for Standard IQ and Symbol Search Substitution Classifications According to Gender.....89

Table 4.22 Comparing Two-Way ANOVAs (IQ Classification X Gender) for Mathematics Calculation Utilizing Standard FSIQ and FSIQss.....90

Table 4.23 Standard FSIQ Classifications Compared with Changes in Classification resulting from Intellectual Abilities FSIQ.....92

Table 4.24 Comparison of IQ Classifications between Standard FSIQ and Intellectual Abilities.....92

Table 4.25 Comparing Two-Way ANOVAs (IQ Classification X Gender) for Word Reading Utilizing Standard FSIQ and IA.....93

Table 4.26 Comparing Two-Way ANOVAs (IQ Classification X Gender) for Mathematics Calculation Utilizing Standard FSIQ and IA.....94

Appendix

Table A1 Reading Achievement Standard Score Means and Standard Deviations by IQ Classification and Gender..143

Table A2 Mathematics Achievement Standard Score Means and Standard Deviations by IQ Classification and Gender..143

Table A3 Reading Achievement Standard Score Means and Standard Deviations by IQ Classification and Age Grouping................................144

Table A4 Mathematics Achievement Standard Score Means and Standard Deviations by IQ Classification and Age Grouping................................144

Table A5 WISC-III Factor Score Means and Standard Deviations for Sample and IQ Classifications..145

Table A6 FDI, Arithmetic, and Digit Span Means and Standard Deviations by IQ Classification..145

Table A7 PSI, Coding, and Symbol Search Means and Standard Deviations by IQ Classification..145

Table A8 Comparison of Reading Achievement Standard Score Means for Standard IQ and Intellectual Abilities Classifications According to Gender..............146

Table A9 Comparison of Mathematics Achievement Standard Score Means for Standard IQ and Intellectual Abilities Classifications According to Gender..............146

List of Figures

Figure 4.1 Student Classifications as a Percentage of Entire Sample (*N*=196)............44

Figure 4.2 Comparison of IQ Classifications and Gender....................................45

Figure 4.3 Histogram of WISC-III FSIQ range in Sample....................................46

Figure 4.4 Comparison of Gender and Age of Participants in Sample........................47

Figure 4.5 Comparison of Gender and Grade of Participants in Sample......................48

Figure 4.6 Average IQ of Males and Females in Sample......................................54

Figure 4.7 Histograms of Age According to FSIQ Classification.............................56

Figure 4.8 Histograms of Grade According to FSIQ Classification...........................57

Figure 4.9 Boxplots of Reading Achievement According to IQ Classification..................60

Figure 4.10 Boxplots of Mathematics Achievement According to IQ Classification.....62

Figure 4.11 Achievement Test used According to IQ Classification...........................67

Figure 4.12 Estimated Marginal Means of Word Reading by IQ Classification and Gender..72

Figure 4.13 Estimated Marginal Means of Math Calculation by IQ Classification and Gender..73

Figure 4.14 Estimated Marginal Means of Reading by IQ Classification and Age Group...76

Figure 4.15 Estimated Marginal Means of Math Calculation by IQ Classification and Age Group...77

Figure 4.16 Significant Differences Between Symbol Search and Coding.....................82

Figure 4.17 Changes According to Classification Resulting from Symbol Search Replacement 84

Figure 4.18 IQ Classifications According to Method of Summarizing WISC-III Results..101

Chapter 1

Introduction

Purpose of Study

This study contributes to the ongoing research in the field of School Psychology by examining some of the effects of using the Full Scale Intelligence Quotient (FSIQ) to classify students aged 6 to 16 years according to their results on an individual measure of intelligence, the Wechsler Intelligence Scale for Children, Third Edition (WISC-III). The subject pool was 196 students classified as Average (IQ ranging from 90 – 109), Low Average (IQ ranging from 80 – 89), or Borderline (IQ ranging from 70 – 79) with data obtained from the Psychological Testing Center of the Education Clinic housed in the Faculty of Education at the University of Alberta.

There were two main objectives in this study. The first objective was to explore the psychometric implications of utilizing IQ-based classifications to differentiate students. The second objective was to compare Average, Low Average, and Borderline students' performance on standardized tests of academic achievement. These measures of reading and mathematics achievement were interpreted in relation to students' age, gender, grade placement, academic history, family composition, and co-morbid learning and behavioral problems in addition to IQ based classifications. Alternative methods of computing WISC-III composite scores or IQ were explored with any resulting changes in classifications reported.

School Psychologists are responsible for translating assessment results into meaningful reports that communicate both psychometric and educationally relevant information. It is intended that this study will provide further insight into the extrapolation of meaning from IQ-based classifications and their utility in grouping children for instructional purposes. Additionally, this paper will question the government-sanctioned use of IQ cutoff scores to allocate special needs funding and provide further support for the 'Rights without Labels' movement.

By combining descriptive statistics and academic achievement results in reading and mathematics with relevant personal and demographic data, a detailed picture of Borderline, Low Average, and Average students emerged. It is widely accepted that interindividual differences exist within WISC-III classifications (Glutting & McDermott, 1994; Ward, Ward, Glutting & Hatt, 1999), but the nature and extent of these differences requires further clarification. If a common goal of intelligence test interpretation is the "classification of individuals according to their cognitive abilities" (Flanagan, McGrew, & Ortiz, 2000, p.14) then it makes sense that the IQ-based classifications should communicate meaningful information. Accurate communication of information is key to advancing both knowledge and understanding of an individual's intellectual abilities and aptitudes as measured by the WISC-III.

Results from this exploratory study may have both direct and indirect implications for all levels of student-centered decision making, from classroom and school, to jurisdictional and governmental. Keeping in mind that evaluation and assessment practices are guided by policies and regulations, it is expected that research in this applied field could influence future policy and guidelines.

Overview

The psychometric measurement of intelligence is controversial (Bidell & Fischer, 1997; Das, Kar, & Parrila, 1996; Devlin, Fienberg, Resnick, & Roeder, 1997; Fischer, Hout, Jankowski, Lucas, Swidler, & Voss, 1996; Sternberg & Grigorenko, 2001b) yet utilitarian (Andrews, Saklofske, & Janzen, 2001; Cooper, 1995; Kaufman, 1994; Meyer et al., 2001; Prifitera, Weiss, & Saklofske, 1998; Sattler, 2001). Rather than adding to the never-ending debate over the measurement of intelligence by attempting to introduce yet another assessment tool or applied theory of brain-behavior relationships, this study suggests a refinement of present practices.

An intelligent approach to assessment demands accountability and change as needed. Informed change occurs incrementally and involves both reflection and action. The current study attempts to combine the disciplines of education and psychology in a manner that may have direct impact on current assessment practices in schools. According to Ackerman, Bowen, Beier and Kanfer (2001) empirical studies must attempt to answer the question "What do we know now that we did not know before the study was completed?" (p. 819). What is known about the epigenesis of intelligence is still open to debate (Bidell & Fischer, 1997; Gardner, Hatch and Torff, 1997; Jensen, 1997), and the field of psychoeducational assessment is wrought with its own set of controversies (Daniel, 2000; Kaufman, 2000; Neisser et al., 1996).

One small step toward refocusing entrenched attitudes surrounding the quantification of intelligence would be to reexamine present assumptions surrounding the measurement and classification of students' intellectual abilities. To this end, several questions needed to be asked. Will Low Average students' WISC-III profiles serve to distinguish them significantly from their Average and Borderline peers? Will patterns of academic achievement emerge that will be distinctive from students' IQ classifications? Is the Full Scale Intelligence Quotient (FSIQ) on the WISC-III the most parsimonious form of comparison, or would alternative summaries of test performance serve to increase validity and reduce the occurrence of false positives and false negatives in establishing psychoeducational diagnoses? Results reported and discussed in later chapters are intended to address these questions within the limits imposed by the archival nature of the data collected.

If a common goal of intelligence test interpretation is the "classification of individuals according to their cognitive abilities" (Flanagan, McGrew, & Ortiz, 2000, p.14) then it makes sense that the resulting classifications communicate meaningful information. The Wechsler Intelligence Scale for Children – Third Edition (WISC-III) employs the Full Scale Intelligence Quotient (FSIQ) to summarize an individual's overall performance and forms the basis for deriving a nomothetic system based on the following seven standard score derived classifications: Very Superior (130 and above), Superior (120 – 129), Above Average (110 – 119), Average (90 – 109), Low Average (80 – 89), Borderline (70 – 79), Intellectual Deficient (69 and below). School Psychologists invariably refer to WISC-III classifications in their psychoeducational reports, and these labels often form the basis for communication with other professionals and nonprofessionals alike.

The present study aims at clarifying this communication process by exploring in depth the WISC-III profiles of a random sample of subjects whose FSIQ has fallen within the targeted 70 to 109 range. A thorough exploration of psychometric similarities and differences between and among Borderline, Low Average and Average students' WISC-III profiles forms the primary impetus for this study. If it can be shown that the differences within classifications are excessive, then the utility of the FSIQ based classification system would be questioned. Several studies have targeted specific clinical and nonclinical groupings of students to confirm the factor structure of the WISC-III (Burton, Sepehri, Hecht, VanderBroek, Ryan, & Drabman, 2001; Roid & Worrall, 1997; Sullivan & Montoya, 1997; Tupa, O'Dougherty-Wright & Fristad, 1997) and its role in consistently identifying various groups of exceptional children (Glutting & McDermott, 1994; Glutting, Youngstrom, Ward, Ward & Hale, 1997; Ward, Ward, Glutting & Hatt, 1999). The psychometric soundness of the WISC-III has been repeatedly demonstrated.

Critics of the psychometric approach to the measurement of intelligence are quick to point out the damaging effects of labeling and mislabeling students (Gardner, 1998; Kronick & Hargis, 1998; Sternberg & Grigorenko, 2001b; Tzuriel, 2001). The present study addressed some of these concerns by exploring patterns of cognitive and academic performance both within and between three separate IQ classifications. As well, alternative methods of calculating and reporting WISC-III results were explored in some depth. Intellectual Abilities (IA) is a term adopted in this study to refer to the General Abilities Index that is determined by combining the Verbal Comprehension and Perceptual Organization indexes from the WISC-III (Weiss, Saklofske, & Prifitera, 1999). The IA is presented as an alternative to the traditional Intelligence Quotient (IQ) summative score. The rationale behind changing the manner in which global WISC-III results are summarized is based on the assumption that when the Verbal Intelligence Quotient (VIQ) and the Performance Intelligence Quotient (PIQ) scores are significantly discrepant, or subtest scatter is extreme, the Verbal Comprehension Index (VCI) and Perceptual Organization Index (POI) may represent purer measures of an individual's overall intellectual functioning. Additionally, the abbreviated format of the IA is a departure from the traditional IQ both in name and in substance. The norm tables for the General Ability Index, or IA, were computed using the Canadian WISC-III standardization sample of 1100 children ages 6 to 16 years.

For several years, Kaufman (1994) has advocated that Symbol Search should be substituted for Coding to calculate the Performance Scale and Full Scale IQs. Symbol Search's higher correlations with these scales, combined with the fact that it is mainly a measure of mental processing as compared to Coding's measure of fine-motor skills, results in a strong case for this routine substitution. Following this recommendation is facilitated by the use of Canadian norms (Saklofske, Hildebrand, Reynolds, & Willson, (1998).

As a result, each participant in the study had three global IQ scores: FSIQ with Coding, FSIQ with Symbol Search, and IA based on the General Ability Index. It was predicted initially that some changes in classification (Average, Low Average, Borderline) would result, and the frequency, type and significance of these changes were analyzed and reported in subsequent chapters. Some significant trends within classifications were found to be dependent upon the method employed in the calculation of the global IQ. Thus, it could be argued that depending on their initial IQ classification, one of these three methods might prove to provide a better representation of the individual's estimated level of cognitive functioning.

Change implies challenging current practices, investigating options, and considering alternatives. Clearly lacking in educational research is a focused attempt to bring theory and practice together in an informative yet scientific manner. One of the main goals of the present study is to fill the gap between assessment of intelligence and the instruction of students. School Psychologists are responsible for using their assessment results to build a solid set of recommendations that can be enacted by parents and professionals alike. Frequently, teachers in the schools are left with the task of utilizing these assessment results to benefit their students without a clear understanding of the academic implications of reported IQ-based classifications. WISC-III results will assist in coding students as developmentally delayed, learning disabled, or gifted, but the benefit that may come with the knowledge that a student is “Low Average” is less obvious

From an initial review of the relevant research literature, there is a noticeable absence of research devoted exclusively to individuals classified as "Low Average" on the Wechsler Scales. As part of the first objective of this study, the psychometric nature of Low Average students, as defined by their performance on the WISC-III, was explored. This involved making both inter-individual and intra-individual comparisons within and across age and sex groupings, as well as comparisons with students from the Average and Borderline classifications. More specifically, in the WISC-III, an analysis of factor scores (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed) was completed. The WISC-III Full Scale Intelligence Quotient with Coding (FSIQcd), with Symbol Search (FSIQss) and Intellectual Abilities (IA) scale were compared to results on measures of academic achievement. Students' results in the areas of reading and mathematics were examined to determine achievement trends according to each of the three methods of determining IQ scores. Additional information was used to further describe the research sample, with attention focusing on variables such as age, gender, grade retention, Special Education coding, and comorbid learning or behavioral conditions.

Typically, when discrepancies between achievement and potential are found it is assumed that the difficulties lie mainly within the student. However, Hargis (1997) maintains that this pattern of underachievement is more often symptomatic of deficiencies in schools. Hargis contends that low achieving students are often viewed as "defective products" who "don't fit into the structures and standards imposed by the schools" (p. 7). Their ability to learn to their potential is not impaired but rather the lock-step nature of school's curricula imposes an artificial timeline that is adhered to so rigidly that "Low Average" students may experience difficulty keeping up the pace. An assumption in this study is that "Low Average" students may have an increased risk for failure due to an incongruance between their academic needs and the availability of appropriate school programming. The first step in establishing the connection between academic aptitude and academic performance would be to identify some of the fundamental ability and achievement variables evident in a sample of "Low Average" students. Similar to their Average or Borderline peers, Low Average students may achieve either within, below or above the low average range in measures of academic achievement. Low Average students may have been excluded from being classified as learning disabled because their IQ's did not fall in a "normal range" (Rourke, 1998). By comparing actual achievement scores in reading and mathematics to IQ classification levels, discrepancies between measures of ability and actual achievement became more obvious. These comparisons added to the descriptive picture that was formed and could lead to conclusions that might question some of the exclusionary practices in special education funding of learning disabled students based on IQ cutoff scores in the Average range.

The "Low Average" classification utilized by the WISC-III presumes a normal distribution with approximately 16% of the population falling within the 80 to 89 standard score boundaries. This percentage indeed constitutes a significant portion of any given student population although, it may be argued that demographics alone could skew a sample such that the actual numbers of "Low Average" individuals may be increased or decreased accordingly. In other words, a sample of students drawn from higher socioeconomic status (SES) schools might have fewer Low Average and more Above Average students than schools from lower SES neighborhoods. No causal link between income and IQ is presumed, although lower SES is associated with an increase in risk factors that might impact children's ability to take advantage of learning opportunities offered in their respective schools. Research in Latin America has clearly demonstrated "highly significant differences in IQ distribution measured through the Wechsler Intelligence Scale for Children (WISC), between children of middle-high socioeconomic

(SE) level and children of middle-low SE level, with significant advantages for the former group" (Rosas, 2004, p. 397). The influence of SES on IQ measures transcends geopolitical boundaries as witnessed by the development of "norm tables for different SES groups in the manual for the Turkish WISC-R" (Gulgoz & Kagitcibasi, 2004, p. 257). International research into the SES and IQ connection repeatedly demonstrates similar results, especially when the relationship between the quality of schooling, poverty levels and psychological factors such as memory and reasoning abilities (Baral & Das, 2004) were investigated.

Beyond Wechsler's classification system, educators use the term "slow learner" was used to refer to those individuals who "can progress in school but cannot grasp higher abstractions and symbols and who advance one-half to three-quarters of a grade per year" (Shelton, 1971, p. 17). In the past it was generally believed that the main cause of the "slow learner's" lack of progress in school was genetic. It has since, however, been shown that the interactive effects of genetics and environment combine to determine an individual's IQ or global intellectual level (Bidell & Fischer, 1997; Sparrow & Davis, 2000).

The "slow learner" label is clearly ambiguous. As Shelton (1971) aptly points out, it is a relative term: "Slower than whom? – Slower in learning what?" (p. 16). School psychologists and educators are charged with the responsibility of answering these questions. By utilizing the normal curve to rank and compare individuals, the psychometric approach to assessment is responsible for creating IQ-based classifications. The present study intends to go beyond previous research that focuses on these groupings as homogeneous in nature. By investigating individual cognitive and achievement profiles, as well as examining alternative methods of summarizing global WISC-III scores, the heterogeneity both within and between the Average, Low Average and Borderline classification was highlighted. Results from this study could lead to increased accountability in the reporting of test results through a de-emphasis on classifications or categorizations of individuals based solely on their psychoeducational assessment results. This in turn could influence the types of recommendations that school psychologists may generate from their psychometric based tests of intelligence.

Labels such as "Low Average" may assist communication between professionals and their clients, yet it begs the question regarding what exactly is being communicated. These "80's kids" may or may not be cognitively and academically similar. Previous research has not fully elucidated the nature of this psychometric grouping, and the present study attempts to undertake this important and relevant task.

In the classroom, "slow learners" may be differentiated from their peers in any number of ways despite their efforts to blend in socially, if not academically. Intra-individual differences in performance may be evident in only one subject area. "Slow learners'" profiles may approach that of students with Learning Disabilities until a closer examination reveals that their academic deficits are not as extreme but more in keeping with predicted levels of performance based on their "Low Average" cognitive abilities. In some cases, "slow learners" may have performance deficits in a number of academic areas. They may be perceived as struggling to keep up with the class and maintain passing grades yet, with persistence and support, they remain within the regular classroom and curricula. When an ecological approach to learning is considered, the question always remains as to whether the specific learning difficulties of the "slow learner" are inherent within this individual or indeed secondary bi-products of insufficient or ineffective accommodation to their learning needs. According to Shelton (1971), once "slow learners" are no longer in school they may not be easily detected in the general population. If this is indeed the case, then the validity of classification systems based strictly on standardized test scores is questionable, if not discriminatory. Kaufman (2000) reminds clinicians that they must use their test results to "generate hypotheses about an individual's assets and deficits and then confirm or deny these hypotheses by exploring multiple sources of evidence" (p. 453). A clear connection between IQ and school achievement has been established, but estimates of lower intellectual abilities need not predict failure. According to Kaufman (2000), test results can translate into successful interventions that will promote learning and improve achievement levels despite less than average performance on measures of intelligence.

Frequently, the reporting of an individual's performance on standardized tests fails to describe adequately the individual completing these tests. Of particular importance is the sociocultural milieu in which the student resides as well as the school culture that is typical of his or her present and past educational experiences (Ogbu & Stern, 2001; Tzuriel, 2001; Vygotsky, 1978). Tzuriel echoes Vygotskian principles when he warns that interpreting a child's IQ as a reflection of their abilities can be misleading when their performance "reflects their entire socioeducational history and is not limited to inner factors" (p. 456). Psychometric assessment employs a host of assumptions based on normative decisions that reflect cultural standards and mores. For the most part, classification decisions based strictly on IQ cutoff scores ignore basic human differences commonly associated with SES and ethnicity. A secondary goal of this study is to examine the heterogeneous nature of "Low Average" students. The efficacy of utilizing IQ-based cutoffs for the purpose of limiting access to specialized programming and government funding could be challenged on ethical and psychometric, if not legal grounds.

The second chapter provides an overview of research literature related to this study. An attempt was made to highlight both controversy and consensus in the assessment field and, more specifically the resulting impact on the role of the school psychologist. A case is made for adjusting present assumptions surrounding the meaning of results obtained from standardized measures of intelligence. Cognitive, neuropsychological, and ecological perspectives are considered. The third chapter delineates the research framework and rationale. Specific research questions and data analysis methodology are outlined. The fourth chapter reports descriptive data obtained from the 196 archived files along with results from inferential statistical procedures utilized to test the null hypotheses of each of the four research questions. Finally, the fifth chapter summarizes these statistical findings as well as relevant information obtained from individual files. Significant trends are reported and the practical significance of these findings to present day assessment practices in the field of school psychology. Additionally, the final chapter outlines some of the study's delimitations.

Chapter 2

Literature Review

The following literature review is divided into four main sections and ends with a summary. In the first section, the political milieu in which many schools psychologists complete their clinical work is considered. Some of the implications of classifying a student as Low Average on the WISC-III are considered along with criteria used to diagnose learning disabilities. The stage is set for the second section that addresses, in more detail, the role of the school psychologist. Their role as problem solver and change agent applies not only to the clients they serve but to the profession itself. The third section provides a brief overview of some of the theories of intelligence and cognition along with a review of the approaches to measurement of IQ. A balanced approach is attempted recognizing that there is no shortage of opinions available when assessment related topics are considered. The fourth section considers future directions of assessment from creative and practical perspectives. The literature review ends with a summary that attempts to highlight significant points that underscore the need for research in the field of assessment that could have direct impact on the school psychologist's role.

Politics of Assessment

When students present with a constellation of complex educational needs it is incumbent on school psychologists to ensure that all levels of their investigation respond appropriately to this complexity. The common ground for all stakeholders in the psychoeducational assessment process is an underlying concern for the welfare of the individuals being assessed.

An inherent conflict or tension between the needs of the individual student and the needs of the educational system is not uncommon. Fulfilling one's academic potential in the school setting may require some tailoring of instructional practice. Individual differences in intellectual abilities do exist. How these differences are measured determines in part how the educational system will respond to the unique learning needs of certain subpopulations of students entrusted to its care. For some time there have been attempts to improve instruction by providing for individual differences: "Matching students with instructional practices conducive to their own heightened productivity and morale is in line with the liberal ideal of a formal education system that is ultimately tailored to the needs of each and every student (McCann, Short, & Stewin, 1991, p. 101). By examining the cognitive and academic variables of a selected population of "Low Average" students, it is hoped that information can be gained that will assist this tailoring process.

Limited resources must be distributed in what is judged to be an equitable manner. Accountability for educational expenditures and their judicial allocation is both desirable and necessary. School psychologists may be misperceived as gatekeepers who control the school's access to special needs funding and augmentative services. Professional practice dictates a strict adherence to the Canadian Code of Ethics for Psychologists, as well as the principles and guidelines as set forth by the Provincial College of Psychologists. Application of these rules and regulations is completed within the context of legislated policy and regulations as they relate to each individual being assessed. A priori decisions regarding the method utilized to calculate an individual's IQ for example will comply with ethical guidelines but may not in fact be the best representation of the individual's cognitive strengths. The present study examined the effects of three methods of calculating an aggregate IQ along with the prevalence of grade retentions, program modifications or adaptations and DSM-IV diagnoses in students classified as Average, Low Average and Borderline. The method used to determine individual IQs not only determines their classification status but may also determine eligibility for special education programming and funding. IQ cutoff scores limit the distribution of government funding and resources but it need not limit expectations of an individual's present and future academic performance.

Within the context of a dynamic school system, what consistently remains unchanged are the preconceived notions of ability or potential that are attached to the results of standardized tests of cognitive ability. Classification implies generalization, therefore, basic logic would dictate that not all characteristics associated with a designated subpopulation of students would necessarily apply to all of its members. This remains the primary impetus for the present study, to explicate the psychometric implications of IQ-based classifications, and through this process, highlight some of the effects of using the WISC-III to classify students as Average, Low Average and Borderline.

School psychologists must continually answer to claims that standardized test are either conceptually flawed or culturally biased. Attempts at quantifying intelligence through traditional psychometric means has lead to several abuses not the least of which has been the reinforcement of racial and ethnic stereotypes (Fisher, Hout, Jankowski, Lucas, Swidler, & Voss, 1996).

It is alleged that standardized test items frequently do not include words or concepts that are familiar to lower SES students (Neill, 1995). Suzuki, Short, Pieterse, and Kugler (2001) agree that higher SES is related to higher scores on intelligence tests and cite additional confounding variables such as level of English knowledge and degree of familiarity with the dominant culture. Although test makers attempt to use various strategies to detect potential test bias while developing their instrument, Susuki et al. (2001) claim that predictive validity may be compromised when these measures either overestimate or underestimate the "true scores" of a particular group. This only adds to the growing public sentiment that standardized tests are unfair and may have limited usefulness when it comes to improving educational access and equality (Neill, 1995; Nettles & Nettles, 1995).

Even when socioeconomic and cultural factors are considered as possible sources of test bias, it is of primary importance to understand fully the goal or purpose of intelligence testing (Puente & Salazar, 1998). This assists in establishing a direction that will guide the assessment process. When the primary goal is to measure intellectual abilities it is necessary to be contextually sensitive to cultural and experiential differences that may impact negatively on the results. An alternate purpose of intellectual testing may be to determine the individual's position within a culturally determined intellectual spectrum. This may be less a measure of innate cognitive abilities but rather an indicator of the individual's ability to adapt and assimilate within the majority culture. In some cases standardized test results have underestimated the cognitive ability of individuals from low socioeconomic levels (Tzuriel, 2001). Those involved in psychoeducational assessment need to remain cognizant of the cultural biases inherent in test construction and the problems related to the transportability of these psychometric instruments across diverse groups within society (Sternberg & Grigorenko, 2001a). The alleged "gatekeeping" function of some forms of standardized assessment (Neill, 1995) can be reduced when these psychometric tools are used judiciously. This assumes that ethical practice will translate into a tailoring of assessment batteries to meet individual needs rather that a "one-test-fits-all" mentality. This tailoring process needs to start at the first level of intervention when the school psychologist administers their battery of standardized tests.

Even with its many inherent flaws, psychometric estimates of global ability, such as WISC-III scores, continue to have a useful role to play in many assessment environments. For example, in the diverse school population there remains a group of students that are often overlooked by teachers and school psychologists alike. They are classified as Low Average by their WISC-III FSIQ results with standard scores ranging from 80 to 89. Many of the students in the Low Average range do experience academic difficulties and may pose a challenge to those professionals who are responsible for instruction and programming decisions.

It needs to be recognized that some of the Low Average students' academic problems are created and sustained by inflexible academic and evaluation standards (Kronick & Hargis, 1998). Willingness to create an Individual Program Plan (IPP), for example, often depends on the teacher's acceptance of its necessity. The student in question might be perceived as creating their own academic problems as opposed to the rigidity of their learning environments. Accommodations are usually made for select groups such students with Learning Disabilities. Presumably they are inherently more capable yet the manifestations of their disability may result in standardized achievement scores that are considerably lower in specific areas than their non-disabled Low Average peers.

Typically Low Average students can not be diagnosed as Learning Disabled due to the fact that a severe discrepancy between ability and achievement is one of the main criteria for categorizing a learning disability (Evans, 1990). This would translate into standard scores on measures of academic achievement in the 50 to 60 range. This study did not attempt to determine the type and extent of ability-achievement discrepancies within the sample due to concerns relating to the validity and reliability of such findings given the number of different achievement measures utilized as well as their American norms versus Canadian norms for the WISC-III. Instead, comparisons of achievement levels of students classified as Average, Low Average and Borderline according to three global IQ measures, FSIQ with Coding (FSIQcd), FSIQ with Symbol Search (FSIQss), and the IA (General Ability Index) were completed. The statistical significance of these findings is reported in the fourth chapter.

Ward, Ward, Glutting, and Hatt's (1999) research on subtypes of learning disabilities (LD) examined ability and achievement profiles on the WISC-III and Wechsler Individual Achievement Test (WIAT) for 201 students identified as LD. Results from a hierarchical cluster analysis using a minimum sum of squares method revealed five distinct clusters or groups. Only two of these groups showed significant discrepancies between ability and achievement in reading and written language. With the remaining three groups, ability levels ranged from the average to low average levels with commensurate achievement in reading and written language. Presumably other variables may have been utilized to facilitate the LD diagnosis but it was obvious that they did not meet discrepancy formula criteria.

What are the academic implications when a Low Average student achieves within predicted levels on a standardized measure of achievement? Using the Low Average student's WISC-III FSIQ, their predicted performance on standardized measures of academic achievement would be expected to fall within a corresponding low average range. When actual and predicted achievement results are compared, significant differences in performance can be determined. What if the Low Average student's achievement results fall within predicted levels based on their age and FSIQ? Typically this student is experiencing academic problems in school yet normative comparisons provide a somewhat different picture. Clearly the Low Average student has demonstrated an ability to learn but at a different rate or pace than the Average student. It is generally accepted that many Above Average students are capable of progressing quickly through their assignments and courses when given the opportunity. A similar flexibility needs to be applied to the Low Average student who may require more time to progress through the same curricula.

> This would recognize that some students might arrive at the same destination if the pace were adjusted to fit them better. Educators must learn to recognize that arriving at the destination is more important than the length of time it takes to reach it. (Hargis, 1997, p. 22)

This study attempted to transcend the politics and philosophical debates surrounding IQ and the measurement of intelligence by adhering to more utilitarian and pragmatic goals. One obvious limitation is that only the WISC-III is being considered with the exclusion of other known and respected measures of cognitive ability. However, it is generally accepted that the WISC-III is widely used in the assessment field (Kaufman, 1994) therefore any suggested changes in assessment practices that may result from this study could have significant implications in the field of school psychology and beyond.

Role of the School Psychologist

School psychologists often position themselves between the vicissitudes of abstract theory and the demands of clinical practice. It may be argued that the school psychologist needs to have their feet planted firmly in both arenas. Ideally, theory would guide practice but reality often paints a less symbiotic picture. Woolfson, Whaling, Stewart and Monsen (2003) suggest that the 'eclectic' nature of the school psychologist's role "may only tacitly reflect the psychological knowledge that they bring to bear to complex and ill-defined problem situations" (p. 283). In other words, real-life problems often require that a variety of perspectives be considered and not just the psychological.

The call for transparency and accountability in a school psychologist's practice has been traced to the "increasingly litigious climate around education and special needs issues" (Woofson et al., 2003, p. 284). As a result, there has been a move toward evidence-based practice (Fox, 2003) even though experienced Educational Psychologists might rather employ their own methods of working with a client. Reflective practitioners acknowledge that every situation is different even though some variables may be shared. Fox points out that School Psychologists may often be guided by their experience as much as the related research when confronted with diverse individual student needs. Assessment related decisions are one area where evidence-based research and daily practice may not align.

The present study attempts to address one of the more salient tasks of a school psychologist, that is, the measurement of an individual's intellectual and cognitive abilities. Accurate reporting of these results includes acknowledgment of measure error as reflected in confidence intervals. Sattler (2001) recommends the reporting of classification levels based on the obtained standard score. This results in a procedure that on one hand acknowledges a range of scores within which the individual's true abilities are believed to exist while at the same time reducing all verbal and nonverbal tasks to one score, the Full Scale Intelligence Quotient (FSIQ) on the Wechsler Intelligence Scale for Children – Third Edition (WISC-III). The manner in which the FSIQ is calculated not only determines its classification level but may limit associated diagnostic decisions. Discrepancy based definitions of learning disabilities (LD) require significant gaps between measures of intellectual ability and academic achievement (Evans, 1990; Glutting and McDermott, 1994; Ward, Ward, Glutting, and Hatt, 1999). The present study explored various methods of determining a global IQ score from WISC-III results and examined these effects on actual levels of academic achievement. When student fall into the Low Average classification they may require some form of assistance to ensure continued success yet not qualify for a LD designation and associated accommodations. The school psychologist is left with the task of recommending appropriate programming for the Low Average student even when special needs funding might not be available to support its implementation.

If role is defined as "the expectation one has about how to carry out one's work" (Peterson, 2001, p.2001) then one must look more closely at the political milieu that school psychologists typically operate in to understand further the nuances of role perception. The relationship between role and context must not be ignored when responding to calls for change. Various levels of government may have a hand in drafting informed legislation surrounding the allocations of limited financial resources for targeted special education populations. Since psychoeducational assessment is commonly viewed as one of the school psychologist's key roles, the result may be a streamlining of their role in many school systems. The school psychologist as technician is non-political and merely applies their psychometric tools and report writing skills to the student eligibility task at hand. Yet at the same time the school psychologist could be viewed as a “gatekeeper” to valuable resources (Dennis, 2004) when the results of their assessment could translate into government funding for students with special needs.

The call for school psychology services to become more comprehensive in nature may result in a shift in role emphasis from a special education assessor to a problem solver and change agent (Braden, DiMarino-Linnen, & Good, 2001; Bramlett, Murphy, Johnson, Wallingford, & Hall, 2002; Peterson, 2001; Watkins, Crosby, & Pearson, 2001). It is difficult to imagine the role of the school psychologist changing significantly unless there is a shift in mindset from one of technician to that of a social broker or interventionist. According to Peterson (2001) the broker attempts to translate their political skills into action by developing plans that will "reconcile opposing interests and be acceptable to enough of the interests so that they can live with a proposal and allow it to be implemented" (p. 298). To orchestrate change school psychologists can not operate in a vacuum created by theory and philosophical allegiances. Instead, they must be prepared to work directly with all stakeholders from a variety of disciplines, perspectives and agendas.

By reshaping the school psychologists' role, individual assessment and consultation may be de-emphasized in favor of the development and evaluation of programs designed to meet the general learning and mental health needs of all students (Braden et al., 2001). There is little consensus in the current research literature surrounding role delimitation. Results from Watkins et al. (2001) survey of administrators and teachers called for an expansion of school psychology services. The assessment role would not be de-emphasized so much as joined by equally important aspects of "special education, counselling, crisis intervention and behavior management" (p. 64).

Dennis (2004) suggests that implementation of a consultation model acknowledges that “Educational Psychologists have specialist knowledge and skills, but that these are brought to bear in a collaborative way” (p.18). As a result they are not viewed so much as an “expert” as an enabler who can be called upon to support ongoing decision making at the school level and beyond. The school psychologist may offer a wider view of issues relating to the special needs of students and provide a “whole school context” (Dennis) when approaching the problem solving process.

In the absence of provincial or national directives, the local political climate or context in which change is expected to occur would be jurisdiction dependent. As an example, Edmonton Public Schools have determined that one of the eligibility criteria for their LD programs is that students have a minimum IQ of 100. Historically, Bryan and Bryan (1986) have questioned the assumption that students with learning disabilities must have an IQ within the average range and recognized the political and budgetary considerations that might exclude Low Average students from this designation. Bryan and Bryan have challenged educators and researchers alike to provide empirical proof of the benefits of such practices by clearly demonstrating that, "knowing that a child received a low IQ score leads to school programs that are better suited to the skills of the child or simply discourages child and teacher alike" (p. 274). The utility of an LD label diminishes when evaluation is based on edumetrics (Bryan & Bryan) rather than standardized scores on tests of intelligence. The ability of individuals to read and understand curriculum materials and meet teacher-determined standards of performance is not only more relevant to the student in question, but likely a more accurate predictor of their success or failure in school than IQ scores or vaguely worded DSM-IV definitions of specific learning disabilities.

The onus is on individual school psychologists to advocate for necessary changes in current assessment and diagnostic practices. Additionally, reasonable caseload sizes need to be maintained so that the counselling and consultative elements can be emphasized along with the demands of assessment, diagnosis and production of professional reports. Bramlett, Murphy, Johnson, Wallingford, and Hall (2002) report that from their 370 survey respondents, the median ratio of school psychologists to students was 1:1500. The National Association of School Psychologists (2000) has recommended a ratio of 1:1000. Unless the desire for an expansion of the school psychologist's role is combined with a corresponding reduction in caseload size to recommended levels, then lasting and effective change will not likely occur.

The implications for institutions responsible for the training of school psychologist are obvious. The enhancement of clinical skills needs to include those areas of expertise that will be in demand when school psychologists are working in the field (Woolfson, Whaling, Stewart & Monsen, 2003). In some respects they may be viewed as generalists in the area of applied educational psychology with sufficient training in assessment, counselling, and consultation (Dennis, 2004). They are often on the front lines of intervention dealing directly with students, families, educators, and administrators. Demands to expand their role must be met with a reality check. Rather than viewing the school psychologist as the panacea for problem solving, often a team-focused approach is more efficacious.

At least in theory, clinical assessment is recognized as representing more than a collection of test scores (Prifitera, Weiss, & Saklofske, 1998; Sattler 2001). In practice this professional standard often remains illusive and difficult to attain.

Intelligence, Cognition and IQ

In the psychometric paradigm, "the fundamental skill or talent critical to human functioning is intelligence" (Fischer, Hout, Jankowski, Lucas, Swidler, & Voss, 1996, p.26). If the psychometricians main concern was to rank individuals according to their perceived intelligence then there was often little consideration for those human qualities that fell outside the narrow parameters of their tests. The fact that these rankings would be plotted comparatively within the hypothesized "bell curve" seemed to add further testament that the individual's general capacity or general intelligence (g) was indeed being measured. Fischer et al. are quick to point out that within this psychometric paradigm there is little or no attention given to such human abilities as, "determination, self-discipline, empathy, creativity, charm, (and) energy" (p. 26).

"IQ (intelligence quotient) refers to a derived score used in many test batteries designed to measure a hypothesized general ability, intelligence" (Lezak, 1995, p.24). Though professionals from a variety of fields claim to understand and use IQ scores as a means of communicating test results, there is little agreement as to the actual meaning of these particular standard scores. Despite the lack of consensus regarding what is meant by intelligence or IQ, they are frequently referred to as being synonymous (Ogbu & Stern, 2001).

Eysenck (1971) responded to the accusation that psychologists were tending to reify intelligence by their attempts to measure it by stating, "Intelligence is not a "thing" but a concept—just as gravitation is a concept …" (p. 46). As a concept, intelligence can be defined operationally, measured psychometrically and expressed numerically in the sound statistical language of standard scores. Accordingly, most definitions of intelligence emphasize such cognitive processes as abstraction and conceptualization. In addition Eysenck recognized the influence of cultural content in IQ tests and suggested that it might be more appropriate to refer to them as measures of education rather than intelligence.

This study restricts its focus to students' result on the WISC-III while fully acknowledging that this imposes a severe limitation on the description of individual's intellectual abilities. The very nature of psychometrics implies that psychological variables such as intelligence will be quantified or measured through carefully designed and validated instruments. The 4-factor model of the WISC-III offers "the most accurate and stable explanation of the data when the Symbol Search subtest was included" (p. 236) according to Grice, Krohn, and Longerquist (1999). There is a history of mixed support for the 3- versus 4-factor structure with noted and influential scholars such as Sattler (1992) making a case for the former. In addition to Grice et al., Roid and Worral (1997) also put their support behind the 4-factor model. Confirmatory factor analysis, with a sizeable Canadian normative sample replicated results from the American WISC-III standardization sample. As a result, Roid and Worrall recommended the continued clinical use of all four factors (Verbal Comprehension, Perceptual Organization, Processing Speed, and Freedom from Distractibility), and called for further research, "to uncover the psychological meaning and underlying dynamics of performance on each factor dimension" (p. 514). In the second research question, this archival study contributes to the ongoing quest for elucidation of the psychometrically confirmed, yet controversial, WISC-III factor structure.

The three forms of IQ derived from the WISC-III remain less psychometrically controversial, perhaps, than its factor structure. Canivez and Watkins (1999) reported that long-term stability coefficients for the Verbal Scale, Performance Scale, and Full Scale were similar for gender, ethnicity and age with the highest stability coefficients evident for the Full Scale Intelligence Quotient. Even with a population of ADHD children, Schwean and Saklofske (1998) were able to confirm the stability of WISC-III scores over a 30 month test-retest time frame. Given the WISC-III's outstanding reliability (Sattler, 2001), Prifitera, Weiss, and Saklofske (1998) still emphasize that, "regardless of the referral question, users of IQ tests need to remember that tests yields information that is part of the diagnostic and decision-making process." (p. 5).

Information gained from a measure of intellectual abilities can be incorporated into a comprehensive developmental assessment. In doing so, an individual's strengths and weaknesses are determined in relation to a constellation of information gained from interviews, observations, and test data (Prifitera & Saklofske, 1998; Sattler, 2001; Sparrow & Davis, 2000).

Even if a balanced perspective is recommended when utilizing IQ based diagnostic labels, this ideal may not be easily achieved. This aspirational goal presumes that these labels are indeed valid and not reflections of cultural or socioeconomic differences. It has been questioned whether individuals with the same IQ have sufficient similarities to warrant similar educational treatment (Tzuriel, 2001). Even Sattler (2001) concedes that diagnosis and classification should not be viewed as ends in themselves. There is, however, an inherent contradiction in using psychometrically based classifications derived from arbitrary IQ cut off scores. A certain degree of homogeneity is presumed within a classification such as "Low Average" yet Sattler (2001) advises that "In developing assessment findings and recommendations, be guided by the child's performance and not be a classification system of arbitrary cut off scores on an intelligence test" p. 26. At the same time we are lead to believe that when properly used, classification systems will assist etiological studies and improve the delivery of needed services. Yet there are some stark reminders that diagnosis and classification can lead to a host of presumptions about an individual that can be potentially misleading, even damaging (Tzuriel, 2001). To avoid misleading interpretations of test results and impending classifications these results must be contextualized within the broader domain of the individual's socioeducational history as well as related sociocultural and socioeconomic factors.

Although overzealous testers and a misinformed public may have perpetuated some of the myths and misunderstandings pertaining to IQ scores (Witt, Elliott, Daly, Gresham & Krammer, 1994), in the research literature there remain several accounts that support the efficacy of this form of investigation into the nature of intellectual functioning and its associated features (Andrew, Saklofske, & Janzen, 2001; Schwean & Saklofske, 1998; Wong, 1999).

Ogbu & Stern (2001) offer an alternative view of intelligence, "From a cross-cultural perspective, intelligence is a cultural system of thought, a cultural or group's repertoire of adaptive intellectual (or cognitive) skills" (p. 7). Within the wider societal environment ecological niches are created where culturally laden tasks or activities generate cognitive problems for members of that society. Within an ecocultural niche, cultural amplifiers of intelligence emerge as tasks that will require and possibly enhance intelligence (Ogbu & Stern).

Ogbu & Stern's socio-cultural views on intellectual adaptations are embodied within Bidell & Fischer's (1997) constructivist view of intellectual development. According to Bidell & Fischer the epigenesis of intelligence is a constructive not a predetermined process.

> The creative construction of new concepts and cognitive skills through acts of hierarchical integration produces new contexts for the participation of both biological and environmental systems, providing a concrete vehicle for their operation in development, and setting the direction and conditions for their activities (p. 236).

In their view, humans determine their own cognitive outcomes as they actively make sense out of the world and build skills to participate in it. Sparrow and Davis (2000) add to Bidell & Fischer's constructivist view by further detailing the types of interactions that will co-occur at various levels of an individual's intellectual development. Sparrow and Davis refer to cognition as

> "the processes whereby individuals acquire knowledge from the environment. Thus the term cognition refers to the highest levels of various mental processes such as perception, memory, abstract thinking and reasoning, and problem solving as well as the more integrative and control processes related to executive functions such as planning, choosing strategies, and the enactment of these strategies." (p. 117).

In general, definitions and theories of cognition and intelligence refer to multiple processes that combine to produce complex cognitive tasks such as problem solving.

Each component of cognition is representative of how subsystems interrelate within this cognitive domain. Therefore, to understand the nature of cognitive functioning it is necessary to comprehend both the performance of individual components as well as their integrated functioning (Das, Kar & Parrila, 1996; Ogbu & Stern, 2001; Sparrow & Davis; Bidell & Fischer, 1997).

One of the influences of cognitive psychology is that it has suggested a reframing of how intelligence is developed. By maintaining that all learning requires thinking, it follows that "activities and socially supported interactions that develop intelligence are virtually indistinguishable from the kinds of instruction that enable students to think critically about subject matter" (Shepard, 1992, p. 326).

Although a complete consensus may be unattainable there is widespread support for the view that "Intelligence is not an inborn, permanent lump in each person's head. Intelligence can be developed to a great extent by opportunities to think and learn" (Shepard, p. 313, 1992). This fits with Bidell & Fisher's (1997) conception of cognitive skills as being active, contextualized and integrated. These skills do not exist in isolation within the person but entail all levels of participation from biological to social. In their view the constructive context of a cognitive skill needs to be taken into account through "research into the mechanisms of cognitive epigenesis and the interrelations among the multiple participating systems" (p. 222). The effects of genetics are context dependent which makes the dichotomous assumptions inherent in the nature-nurture conceptualizations of intellectual development both reductionistic and shortsighted.

Clearly, the manner in which intelligence is defined will predetermine its measurement. This "knotty problem" of "knowing about knowing" (Das, Kar, & Parilla, 1996) serves to mystify many. The evolution of consciousness in representing true knowledge can be viewed from the dualism inherent in top-down and bottom-up cognitive processing models. The bottom-up explanation suggests that symbolic systems used to represent experience and direct behavior evolve through a process of assimilation and integration of external and internal stimuli. As a result "consciousness with discriminating intelligence" (Das et al., 1996, p. 24) develops. The reverse is true with top-down processing where such mental functions as "reasoning, motivation, and the sense of self" (Das et al., p. 24) facilitate the representation and integration of experience.

If the integration of internal and external sensory stimuli and human cognition are inextricably linked to intellectual functioning, then the construction of "culture free" tests of intelligence may indeed be an impossibility (Eysenck, 1971).

The "cognitive revolution" has been attributed with forging a new direction in the study of intelligence and its measurement. On the forefront of this emerging perspective Das, Kar, and Parrila (1996) have combined both theory and practice to outline an alternate course to the traditional psychometric paradigm.

Human cognitive processes were described within the framework of three functional units. "Luria's work on the functional aspects of brain structures formed the basis of the PASS model and was used as a blueprint for defining the important components of human intellectual competence" (Das, Kar, & Parrila, 1996, p. 49). Blending together the neuropsychological, cognitive, and psychometric approaches the PASS model operationally defines human intellectual competence and the architecture that presumably underlies its processing.

Briefly the PASS model encompasses three interactive components: Planning, Arousal/Attention, Simultaneous and Successive processing. These components operate within the context of an individual's knowledge base receiving both serial and concurrent input and producing output that is either serial or concurrent in nature. Understanding the role of knowledge is central to fully appreciating the ecological nature of human intellectual development.

Das (1998) questions if intelligence is synonymous with general mental ability when he suggests that the definition could be broadened to include the general characteristics of a person. To corroborate further this contention he states, "Some people have very little practical intelligence, although they might be great scholars. At the same time, we find people who have done well in business, and who have an extremely good practical sense, but are poor intellectuals" (p. 205). However, intelligence is more than a collection of abilities and talents. Das acknowledges that the cognitive activities of knowing, understanding and thinking are subsumed under this expansive concept. The complexities of cerebral interconnections at the neuronal level serve to underscore the monumental task of deciphering the origins of thoughts, feelings and actions that psychologists attempt to define and delimit as signs of intelligent behavior. To assume that any test of intelligence could possibly summarize entirely the mysterious workings of the human mind borders on fallacy, if not folly.

General ability or 'g' has no locus within the brain. Instead it was considered to reflect the individual's ability to perceive associations across neuropsychological processes and subprocesses. General ability was thought to be effected by dysfunctional processes within the brain and thus sensitive to its intactness. "Neuropsychological studies demonstrated that there is no general cognitive or intellectual function, but rather many discrete ones that work together so smoothly when the brain is intact that cognition is experienced as a single, seamless attribute" (Lezak, 1995, p.23). It is believed that functional magnetic resonance imaging will continue to reveal more about the locus of mental subprocesses assessed on tests of intelligence (Prifitera & Saklofske, 1998).

Psychometricians have been accused of defining intelligence "after the fact." In other words, their tests were constructed in the absence of explanatory theory and the subsequent results were used to carve out interpretative conclusions as to what indeed was being measured. Thus, the measuring tools themselves define intelligence in the absence of "observations of intelligence at work" (Fisher et al., p. 30). According to Fisher et al., the first psychometricians chose to use the normal curve to represent the distribution of intelligence based on assumption not fact. Also, from a technical standpoint it was easier to convert test scores using a bell-curve distribution than other possible statistical distributions.

Sattler (2001) provides an excellent summary of some of the more salient definitions of intelligence and concludes that there are six major interdependent aspects that are common to many of them. His list includes the following: "knowledge-based thinking; apprehension; adaptive purposeful striving; fluid-analytic reasoning; mental playfulness; and idiosyncratic learning" (p. 137). How you define this abstract concept has direct implications on how you might attempt to quantify an individual's intellectual abilities. On one hand it would seem to be an impossible task yet there does appear to be some agreement in the research community that intelligence tests do adequately measure at least some of the more important aspects of intelligence (Sattler, 2001). This research study attempts to take these beliefs one step further by providing an assumption check, that is, to examine more closely some of the methods of calculating IQ and determining whether these IQ's are adequate predictors of achievement.

School Psychologists continue to respond to the demands for psychometric testing and the determination of IQ's despite growing concerns that funding eligibility is being ascertained rather than a measure of students' intelligence.

Future Considerations

Throughout much of the past century the study of intelligence was carried out by psychometricians who were mainly interested in the measurement of abilities and traits (Gardner, 1998). In Gardner's opinion this resulted in a scientifically conservative perspective that emphasized a unitary rather than pluralistic explanations for individual differences in intelligence. "Fortunately, over the last two decades, the study of intelligence has become much more varied in terms of definitions, approaches, sources of data, and tentative conclusions" (Gardner, 1998, p. 1).

Gardner himself provides a prime example of creative intelligence in action. Surveying evidence from a range of sciences, his initial list of intelligences included seven: language, logic, spatial abilities, musical, bodily, interpersonal and intrapersonal. Gardner believes there is enough evidence to support an eighth intelligence, naturalist's intelligence and is looking at the possibility of a ninth, or existential intelligence. His definitions of intelligences are not based on test correlations. Taking more of a qualitative approach his research synthesizes information gained from a variety of biological, cultural, and psychological sources.

Purposefully, Gardner (1998) has not developed a battery of tests for these different intelligences. Assessments of multiple intelligences would need to satisfy two criteria, "First, they need to be direct and as natural as possible – the assessment should not occur through a paper-and-pencil instrument. Second, they need to survey an intelligence in some detail" (p. 2). This form of assessment is similar to the ecological approach (Witt et al., 1994) that examines the complexities of interrelationships between a variety of stakeholders. There are also some elements of functional or authentic assessment when the mastery of specific tasks or abilities is considered to be of prime importance.

As interesting as Gardner's theories may be, psychologists are not prepared to completely abandon their reliance on standardized tests. In recent years theory based measures of intelligence have emerged providing a perspective that differs from the traditional psychometric model (Das et al., 1996). Brain-behavior relationships are continuing to be mapped out with future research destined to explain its intricacies. The controlled actions of numerous genes have been associated with learning and memory (Wahlsten & Gottlieb, 1997). In the opinion of Wahlsten and Gottlieb, "There is no longer any refuge for doubt on this question. Environment regulates the actions of genes, and genes, via changes in the nervous systems, influence the sensitivity of an organism to changes in the environment" (p.178). The interdependency of environment and genotype has been established yet not fully understood.

In the words of Robert Sternberg (1998),

> The study of intelligence is like a real world Jeopardy game. Curiously, there is more agreement regarding answers than there is regarding what questions these answers answer. ... To understand the field of human abilities and intelligence, one must consider questions at least as much as answers" (p. 1).

Sternberg reminds his readers that the paradigms one uses to understand human abilities and intelligence will influence the questions generated. Researchers are challenged to explore historical and international perspectives before pursuing isolated studies that fit only within their limited perspectives.

Flanagan, McGrew and Ortiz (2000) state that there is a need for intelligence testing to merge with intelligent interpretation. They offer the G*f* - G*c* theoretical model as an empirically researched alternative to previous attempts at establishing construct validity for the Wechsler Scales. "It is clear that meaningful use and interpretation of the Wechsler Scales require the adoption of an alternative (4th wave) approach in which contemporary theory, research, measurement principles and hypothesis validation are integrated" (p. 13).

The G*f* - G*c* theoretical model referred to by Flanagan, McGrew and Ortiz (2000) goes beyond a simple dichotomous conceptualization of human cognitive ability based merely on Fluid (G*f*) and Crystallized (G*c*) Intelligences. In discussing G*f* - G*c* theory Flanagan et al. (2000) refer to *g* or cognitive abilities as being classified as either broad (stratum II) or narrow (stratum I). The hierarchical structure of G*f* - G*c* theory therefore subsumes a taxonomy of multiple intelligences within a psychometric framework. Thus Flanagan et al. have attempted to create both an empirical and theory-based approach for guiding interpretation of the Wechsler Scales.

Summary

Restrictive government funding for financing special education programming seems to demand the classification of students. School personnel scramble to meet these ever-changing regulations and coding guidelines in the hopes of better serving those students who have been judged to have more intense learning and behavioral needs. The school psychologist is often called on to be the hapless pawn in this chess game of funding and special service allocation.

Rather than perpetuating an ill-conceived or inequitable system of assessment and classification, taking a personal stance requires more than a gut feeling that there may be something wrong with these evaluative systems. When a student is classified as Low Average by the WISC-III Full Scale IQ this should not automatically exclude them from needed services or limit their academic options. The method used to calculate an aggregate WISC-III score is a significant factor in determining who meets predetermined IQ cutoffs.

According to Sternberg and Grigorenko (2001a) "Intelligence is a term that is sometimes used to refer to those abilities deemed most important by a society and thus worthy of measurement" (p. 344). Although theory directs practice, there are many competing forces that shape and influence the school psychologist's approach to the measurement of intelligence. Just as the nature versus nurture debate may no longer be adequate in explaining the interactive dimensions of the development of human intelligence, theorists and practitioners that limit themselves to narrow views of assessment may also be in danger of extinction.

The fact that the Wechsler family of tests has survived despite mounting opposition to its psychometric conceptualization of intelligence is a testament to its tenacity if not veracity. Bordering on the brink of being over-studied, there is some consolation in knowing that using it is supported by studies that attest to both its utility and reliability. David Wechsler followed a pragmatic approach in test construction with his primary motivation being to create an efficient tool for clinical purposes. He acknowledged that how intelligence is defined shapes its measurement. Wechsler (1974) defended intelligence testing by recognizing that the way IQ scores are interpreted and used are as much cause for concern as the theoretical underpinnings from which it originates. This research adopted a similar stance in that alternative methods of determining the IQ (FSIQcd, FSIQss, General Ability Index or IA) resulted in some changes in classification (Average, Low Average, Borderline) with some classifications more effected then others depending on the method employed.

Flanagan et al. (2000) believe that the Wechsler Scales' popularity may be attributable to their atheoretical features. Current research efforts seek to go beyond simple pragmatics by examining the salient role that theory plays in test construction and validation (Das, Kar, & Parrila, 1996).

Much like the research questions about how much of a given behavior is due to genetics or environment may be misleading (Bidell & Fischer, 1997) research in educational psychology needs to avoid the trap of continually evaluating the assessment instruments employed in the study of an individual's intellectual functioning.

Openness implies exploration, which in turn demands a rigorous review of related data and outcomes. "Many psychologists contend that intelligence involves the ability to learn and change adaptively in new situations, although they are still searching for some way to separate capacity for learning from amount that has already been learned" (Wahlsten & Gottlieb, 1997, p. 164). A review of current theories of intelligence and cognition support the need for an examination of several factors when considering such complex issues. Genetics, environment, and culture combine in unique and individual ways to enhance or inhibit intellectual growth.

Janzen and Saklofske (1991) remind researchers and educators alike that by improving the quality and quantity of information gained through assessment more effective decisions can be made at all levels. Thus it is necessary to continually monitor the decisions that are being made to ensure that they are having the expected and desired results.

One of the unique contributions of this study is that it is linked closely to the traditional role of school psychologists as they approach the task of measuring students' intellectual and academic abilities. Extrapolation of results from this study may directly affect the methods of summarizing WISC-III test results. In turn diagnosis of exceptionalities, such as Learning Disabilities may be expanded to include Low Average students if sufficient evidence can be present to support the need for change in current exclusionary practices.

The main focus of this study was the description of some of the cognitive and academic abilities of groups of students classified as Average, Low Average, and Borderline by their performance on the WISC-III. Interindividual variation within each classification was found along with changes in IQ-based classifications depending on the method used to determine the Full Scale IQ. A thorough analysis of individual test performance occurred at the ipsative and group levels within the context of the subject's gender, age, grade, and evidence of comorbid learning, attentional and behavioral problems.

Chapter 3

The Study

Research Framework and Rationale

There is a segment of the school population that is often overlooked and under studied by educators and researchers alike. These individuals may be at-risk for school failure even though Special Education services are available within school systems. School Psychologists and School Personnel have a common goal to identify students' needs and match them with available programs and services. However, problems arise when students do not meet the provincial and local eligibility criteria for these specialized programs and services based on their psychoeducational assessment results. Assumptions pertaining to students' academic capabilities stem, in large part, from their IQ based labels and classifications. The present study attempts to question these assumptions and search for evidence that will serve to verify or refute claims that Low Average students differ significantly from students in neighbouring classifications.

The population of students studied is defined by the Wechsler Intelligence Scale for Children - Third Edition (WISC-III) in terms of their Full Scale Intelligence Quotient (FSIQ). They are classified as "Low Average" by their FSIQ with results falling between the standard scores of 80 and 89. The objectives of the study are to examine the cognitive profiles of this group, highlighting observed strengths and weaknesses in combination with their levels of word recognition and mathematics performance on standardized measures of achievement. Comparisons between the three groups (Average, Low Average, Borderline) were made for each of the three methods of calculating the FSIQ. This was compared to their actual levels of academic performance in order to determine if significant differences existed between the three classification groups. Additionally, the four WISC-III factor scores (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, Processing Speed) were compared across IQ-based classifications according to the standard FSIQcd method of calculating IQ.

When the FSIQ is at either the upper or lower ends of this ten-point range and measurement error is taken into account, the Low Average student would not differ significantly from those individuals classified as Borderline or Average respectively. Conversely when confidence intervals are considered, students classified as Average or

Borderline may have I.Q.'s falling into the Low Average range. When students are classified according to their FSIQ the ensuing label is frequently referred to in psychoeducational reports and in turn may be used as a means of grouping students for instruction in schools. It is possible that more similarities than differences may result when individuals with Low Average WISC-III profiles are compared with Average and Borderline students' profiles. This remains one of the primary goals of the present research, to explore both the interindividual and ipsative cognitive profiles formulated within the psychometric confines of the WISC-III. Given that IQ scores are used as a means of determining eligibility for specialized programs and services it is incumbent upon School Psychologists to interpret their assessment results in a fair and ethical manner. Classifying a student as "Low Average" may automatically exclude them from such designations as Learning Disabled and relegate them to alternative courses of study.

Utilizing the WISC-III FSIQ based classifications system as a means of grouping and comparing participants is a departure from previous research that focuses on standard deviations from the mean to determine subject groupings (Masi, Marcheschi, & Pfanner, 1998). The link between research and practice is more obvious when subjects remained grouped according to their Average, Low Average or Borderline status. It is hoped that the extrapolation of these research findings might lead to direct statements or recommendations that could influence policy and guidelines presently excluding Low Average students from needed programs or services. More broadly, it is possible that utilizing a practice based approach to research will improve the face validity of its results and make the findings more accessible to a wider audience, including teachers and parents as well as academics and the targeted "Low Average" population itself.

Research Hypotheses

The main goal of the this study is to determine the nature of students classified as "Low Average" by a standardized measure of cognitive ability and academic aptitude, the WISC-III Full Scale Intelligence Quotient (FSIQ). The efficacy of using WISC-III standard scores as a basis of classifying Low Average students may be questioned if it can be shown that significant differences in cognitive and achievement profiles exist within the Low Average classification. IQ classifications derived from WISC-III results may not accurately reflect Low Average students' academic abilities or intellectual potential. Alternative methods of summarizing WISC-III test results, as outlined in this study, may alter perceptions related to the immutability of IQ measurement as well as some of the more obvious statistical effects of classifying individuals according to their estimated level of performance on the WISC-III. Nomothetic classifications carry a host of stated and unstated assumptions. The present study strives to improve the transparency of some of the more obvious assumptions that apply to IQ-based classifications, in the hopes that school psychologist might utilize results from intelligence tests more effectively and humanely.

The following research hypotheses served as the main outline for subsequent inferential statistical analysis:

Hypothesis One.

(a) There will be no difference in performance between individuals on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys.

(b) There will be no difference in performance between individuals on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline individuals grouped as Younger (ages 6, 7, 8, 9 years), Middle (ages 10, 11, 12) and Older (ages 13, 14, 15, 16).

Hypothesis Two.

The null hypothesis assumes there will be no significant difference in mean factor indexes - Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), Freedom from Distractibility Index (FDI), Processing Speed Index (FDI) – when results from Average, Low Average and Borderline students are compared. The alternate hypothesis states that significant differences will emerge when IQ classifications and corresponding WISC-III factor scores are compared.

Hypothesis Three.

The null hypothesis indicates that when Symbol Search is substituted for Coding to determine the Full Scale IQ (FSIQ) there will be no difference in mean levels of performances on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys. The alternate hypothesis states that IQ classification resulting from the FSIQ with Symbol Search will differ significantly when mean reading and mathematics achievement levels are compared.

Hypothesis Four.

The Null Hypothesis assumes there will be no difference in mean levels of performances on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys when the Intellectual Abilities (IA), an abbreviated method of summarizing WISC-III results, is employed. The alternative hypothesis states that significant differences between the 3 IA-based classifications will exist when measures of reading and mathematics achievement are compared.

Research Methodology

Testing Hypotheses One (a) and (b), Three and Four.

A 3 X 2 between subjects factorial design was employed to test the significance of IQ-based classification (Average, Low Average, Borderline) and gender on corresponding levels of academic achievement. Reading (Word Recognition) and Mathematics (Calculation) results will be considered separately to reduce the number of possible interaction effects. Similarly, a 3 X 3 between subjects factorial design was utilized to test the significance of IQ-based classification and age groupings on achievement levels for Hypothesis One (b).

It is recognized that the unbalanced factorial design utilized in the two-way ANOVAs relies on unweighted mean to give equal weights to the different treatment groups in determining treatment effects even though a different number of observation may contribute to each treatment mean (Keppel & Zedeck, 1989). However, confidence in the resulting statistical findings can be gained by recognizing that the Type III sum of squares utilized by the data analysis program remains invariant to cell frequencies and "can be used with both balanced and unbalanced designs" (Field, 2000, p. 312).

The WISC-III Full Scale Intelligence Quotient (FSIQ) was calculated according to three methods. The first research question utilized the standard form with Canadian norms as outline in the manual for this instrument. The FSIQ in the third research question employed the Symbol Search substitution method to calculate this global score. This procedure was consistent with Kaufman's (1994) recommendation that the FSIQ should be calculated using Symbol Search rather than Coding as one of the ten aggregate subtests. The fourth research question utilized the IA (General Ability Index) obtained from an 8 as opposed to 10-subtest composite to calculate the FSIQ. Both Coding and Symbol Search are excluded from this aggregate score, effectively forming a third predictor variable. The General Ability Index or IA also excludes the Arithmetic subtest common to the FSIQcd and FSIQss predictor variables. Even though the emphasis on the Arithmetic subtest is not on mathematical knowledge but more on mental computation and concentration, including this subtest in a predictor variable could be considered a confound when using performance on a test of mathematics abilities as one of the criterion variables. This will be discussed further in the following chapter. Resulting changes in IQ classification were noted for the three comparison groups (Average, Low Average, and Borderline) across the three methods of calculating the FSIQ.

Kaufman (1994) "recommended strongly that WISC-III examiners routinely substitute Symbol Search for Coding as part of the regular battery, and use Symbol Search to compute Performance IQ and Full Scale IQ" (p. 60). The following psychometric reasons are provided for this change: Symbol Search correlates more highly with Performance Scale (Symbol Search — .58 across ages; Coding — .32); higher correlation with Full Scale (Symbol Search — .56; Coding — .33); higher loadings on Perceptual Organization Factor (Symbol Search — .54; Coding — .39); and higher *g* loadings (Symbol Search — .56; Coding — .41). Following Kaufman's recommendation is facilitated by the availability of Canadian norms provided by Saklofske, Hildebrand, Reynolds & Willson (1998).

Canadian norms for the General Abilities Index (Weiss, Saklofske, Prifitera, Chen & Hildebrand 1999) contributed positively to the validity of comparing subsequent FSIQ versus General Ability Index (IA) differences. Direct manipulation of the predictor variables can lead to the identification and clarification of causal relationships although the existing associative confounds of IQ and achievement provided some limitations when analyzing these results.

Weiss, Saklofske, Prifitera, Chen, and Hildebrand (1999) established that the "General Ability Index is an appropriate substitute for the Full Scale IQ (FSIQ) score if the Verbal Comprehension Index and Perceptual Organization Index are considered better estimates of the child's verbal and nonverbal reasoning abilities that the Verbal IQ and the Performance IQ scores" (p. 1). This would occur when the Arithmetic subtest on the Verbal Scale is significantly different than the mean of the Verbal subtest scaled scores or when Coding and Symbol Search are significantly different from the mean of the Performance subtest scaled scores. Significant subtest strengths or weaknesses would positively or negatively skew the Verbal and Performance Scale IQs and exaggerate any existing discrepancies. With correlation between the FSIQ WISC-III and the General Ability Index at .96 (Weiss, Saklofske, Prifitera, Chen, and Hildebrand, 1999) this 8-subtest alternative represents a more parsimonious approach to estimating an individual's intellectual abilities.

A basic prediction of this research is that students with a higher FSIQ (predictor variable) will perform better on a standardized measure of achievement (criterion variable) than students with lower FSIQ's. A Two-Tailed *F* Test was used to test this Hypothesis of Difference. The active manipulation of the aggregate IQ score through the number and type of subtests used serves as reminder that measures of intelligent are not infallible. Tradition sometimes replaces reason when it comes to assessment practices. Although there are limits to the amount and type of psychometric manipulations available there is a need to determine the effects of each and compare these results from both a statistical and practical perspective. It is hoped that this study will contribute to growing demands for accountability as well as increased utility of psychoeducational assessments.

Different standardized tests will provide academic achievement information and therefore introduce another source of measurement error. This obvious delimitation could not be avoided given the individual nature of each assessment battery included in the study. The utilization of only the Word Recognition and Mathematics calculation tasks might serve to limit some validity concerns given the similar nature of these measures across the various tests included in the study. This perceived limitation might in fact provide a more accurate reflection of school psychologists' actual work in the field. Typically when they are called to consult with school personnel, student test information is provided by many different instruments. Adherence to standardized testing procedures helps to ensure the reliability of these test results. In turn the standard scores obtained (mean = 100, standard deviation = 15) can be used to make informed comparisons. Unlike age and grade equivalents, each standard score represents "how far an examinee's score lies from the mean of a distribution in terms of the standard deviation" (Sattler, 2001, p. 97). School Psychologists typically refer to students' school records to look for trends in test performance as well as variations in IQ scores over time.

Testing Hypotheses Two

In comparison to the previous research hypotheses the second hypothesis represented a relatively simple comparisons of the 4 factor indexes (VCI, POI, PSI, FDI) of the WISC-III between the 3 IQ classifications (Average, Low Average, Borderline). To avoid possible interaction effects likely if more complex statistical comparisons were performed, the one-way ANOVA was utilized and would adequately test for significant differences between these classifications. The robust nature of this procedure adds to the reliability of the results given the fact that some small differences in sample sizes existed. When significant differences were found between classifications, appropriate post hoc comparisons were performed to determine if this trend was evident with all 4 factor indexes.

Additionally, the full sample was divided according to gender and the same one-way ANOVA procedure was repeated. Resulting classification groupings were smaller but adequate to reliably test for differences in mean factor scores across the 3 IQ classifications.

Analysis of Data

In the behavioral sciences one of the more frequently used statistical methods is the calculation of the relationship between two variables according to the Pearson product-moment correlation coefficient, *r*. In the present study a presumed linear relationship exists between the dependent or criterion variable Y (academic achievement) and the independent or predictor variable X (IQ). r^2 represents the proportion of variance present in either of the X or Y variable, which may be accounted for, by the variance in the other. As part of the data exploration prior to the use of inferential statistics the frequency distributions of such descriptive variables as age, gender, grade, FSIQ, and IQ-based classification were plotted using histograms and frequency polygons. Sprinthall (1994) states, "The frequency polygon is especially useful for portraying two or more distributions simultaneously. It allows visual comparisons to be made and gives a quick clue as to whether the distributions are representative of the same population" (p. 23). All forms of exploratory analysis were supportive of the fact that data gathered from the present sample approached that of a population with a normal distribution. Therefore an Analysis of Variance (ANOVA) technique was used to analyze differences between and among the various age, gender and IQ grouping within the sample.

Hypothesis One.

(a) The WISC-III FSIQ was used to separate individuals into one of three classifications (Average, Low Average and Borderline). The null hypothesis assumes there will be no difference in performance between individuals on two measures of academic achievement (word recognition and mathematics calculation) among girls and boys.

(b) The research sample was further divided into three age grouping to examine the possible effect of maturation on levels of achievement and IQ classification. The null hypothesis states that there will be no difference in performance between individuals on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline individuals grouped as Younger (ages 6, 7, 8, 9 years), Middle (ages 10, 11, 12) and Older (ages 13, 14, 15, 16).

In both parts of the first research question the two-way ANOVAs were completed twice. Once with results from all six achievement test grouped together and a second time with only the Wechsler Individual Achievement Test (WIAT) results. Comparisons were made between subsequent findings to help determine if the grouping of various measures of academic achievement might impact the results. Significant findings were found with the full sample and WIAT only subsample with post hoc results revealing the same trends as well. A detailed accounting of these results is outlined in the fourth chapter.

The determination of age groupings for the second part of the first research question were a product of two considerations, individual developmental levels and equalizing the numbers within each group to increase the strength of statistical comparisons. Subsequently the three age groups (Younger, Middle, Older) paralleled the three developmental divisions common to Canadian school systems (Divisions I, II, III) with a few obvious age-related exceptions. The fact that 20% of the sample experienced grade retentions meant that with several individuals, their age exceeded what might be expected for their grade placement when their psychoeducational assessment was completed. Due to the tenuous and somewhat arbitrary nature of these age groupings, this was the only research question where they were included in the analysis. Consequently, age-related differences were not the main focus of attention in the results discussion that followed in the final chapter.

Hypothesis Two.

> Individual differences in factor indexes - Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), Freedom from Distractibility Index (FDI), Processing Speed Index (FDI) – were examined to determine if Average, Low Average and Borderline students could be differentiated. The null hypothesis states that there is, in effect, no difference between IQ classifications when corresponding WISC-III factor scores are compared.

The main focus of the second research question was to compare the four WISC-III factor indices across the three IQ-based classifications. The Average, Low Average and Borderline groupings were based on the standard method of calculating the FSIQ. A simple one-way ANOVA was used to avoid the complications resulting from possible interaction effects commonly encountered in more advanced statistical techniques. Additionally, performing multiple comparisons with pairs of factor indices, for example, VCI and FDI or POI and PSI, might be considered theoretically unsound given the known associations that already exists between subtests comprising the Verbal Scale (VCI, FDI) and Performance Scale (POI, PSI) factor indices.

Hypothesis Three.

> This portion of the study introduced a second method of calculating the Full Scale IQ (FSIQ) by substituting Symbol Search for Coding. The Null Hypothesis assumes there will be no difference in mean levels of performance on two measures of academic achievement (word recognition and

mathematics calculation) among Average, Low Average and Borderline girls and boys when the Symbol Search substitution (FSIQss) method is employed.

Hypothesis Four.

The fourth element of this study examined the effect of using an abbreviated IQ on subsequent classifications. Intellectual Abilities (IA) instead of FSIQ was utilized to summarize WISC-III results. The Null Hypothesis assumes there will be no difference in mean levels of performance on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys when the IA, an abbreviated method of summarizing WISC-III results, is employed.

The same two-way ANOVA procedure performed in research question one (a) was repeated with the third and fourth research questions. The only difference was the method of calculating the FSIQ which did result in some increases and decreases in individual IQs. As a result two additional IQ classifications appeared, High Average and Extremely Low. Given the limited number in these classifications they were not included in the analysis. As a result the actual size of the Average, Low Average, and Borderline groups change slightly as compared to their corresponding groupings in the first research question. This slight change in sample size was not considered to be a significant confounding variable given the robust nature of the ANOVA procedure and the ample number of individuals being compared.

Wechsler Intelligence Scale for Children - Third Edition (WISC-III)

The WISC-III evaluates abilities among children aged 6 years 0 months through 16 years 11 months. The battery consists of 13 subtests with 10 of these being mandatory. These 10 subtests contribute to two IQ scales: the Verbal Scale IQ (VIQ), which is composed of five subtests, and the Performance Scale IQ (PIQ), which encompasses the other five subtests. The VIQ and PIQ are combined to form the Full Scale IQ (FSIQ). All three WISC-III IQs have means of 100 and standard deviations of 15. Predictors or independent variables for this study are the VIQ, PIQ, and FSIQ.

The WISC-III was renormed with a sample of 1100 Canadian children that included 100 children (50 female and 50 males) in each of the eleven age groupings. This sample was stratified on the basis of information obtained from the 1986 Census

Canada survey according to ethic origin, geographic region, and parent education level.

Average reliability coefficients for the eleven age groupings were reported in the WISC-III Canadian Supplement Manual as follows: FSIQ = .95, Verbal Scale = .93, Performance Scale = .89, Verbal Comprehension Index = .93, Perceptual Organization Index = .88, Freedom from Distractibility Index = .85, Processing Speed Index = .86. These stratified split-half reliability coefficients reflect a high degree of consistency across age grouping for the WISC-III. In this study, groups of Average, Low Average and Borderline individuals were compared. The fact that various combinations of age groupings are reflected within each IQ classification should not have a negative effect on subsequent statistical analysis given these high reliability coefficients.

Chapter 4

Results

Overview

This chapter begins with a general outline of the research sample, including both descriptive information and exploratory statistics. Following this detailed overview a more focused analysis of results will proceed according to the four research hypotheses within the three targeted IQ groupings (Average, Low Average, and Borderline). The chapter concludes with a summary of the parametric statistics along with germane comparisons between the IQ-based classifications, and possible implications of the study that are explored in greater detail in the final chapter.

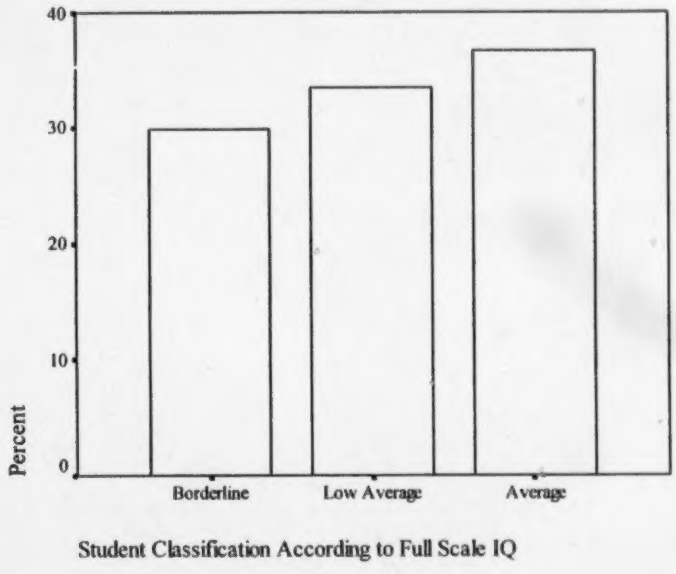
The Research Sample

The main criteria for selecting cases from the archives of the Education Clinic's assessment files were as follows: Wechsler Intelligence Scale for Children, Third Edition (WISC-III) Full Scale Intelligence Quotient (FSIQ) falling into one of three targeted classifications (Average, 90 – 109; Low Average, 80 – 89; Borderline, 70 – 79) and a standardized measure of academic achievement in the areas of word recognition and mathematics calculation. Figure 4.1 illustrates that even though the actual numbers with classifications varied (Average = 69, Low Average= 68, Borderline = 58) these differences might be considered less significant when their percentage relative to the entire sample population is considered.

Cases where statistically significant differences between Verbal Intelligence Quotient (VIQ) and Performance Intelligence Quotient (PIQ) were not included in the sample. When the VIQ and PIQ differ significantly resulting FSIQs are usually difficult to interpret given the wide variability in cognitive strengths and weaknesses evident within these types of WISC-III profiles (Sattler, 2001). Consequently classifications based on any significant VIQ-PIQ 'splits' would have become a confounding variable given their somewhat controversial nature. In other words, the validity of resulting IQ classifications could be questioned especially when compared to the remaining Average, Low Average and Borderline groups where no significant differences between VIQ and PIQ were evident. It is difficult to estimate the total number of cases rejected due to this one factor alone but it is safe to say that it served to reduce the number of eligible cases by a significant amount. A total of fifty-six cases, 30 male and 26 female, were gathered before the decision was made to not include them in the study. Another criterion that affected file selection was the need to have all 10 core subtests as well as two supplementary subtests (Symbol Search and Digit Span) completed. Therefore cases lacking all 4 factor indices were eliminated from the study. Additionally, attempts were made to keep the numbers within each IQ classification relatively equal. As a result, several Average male and females cases were not selected. Following a through review of archived clinic files from 1994 to 2003, a total of 196 cases were found to meet the specified criteria. Figure 4.1 provides a view of the distribution of Borderline, Low Average and Average individuals included in the sample followed by Figure 4.2 that provides a further breakdown of IQ classification and gender.

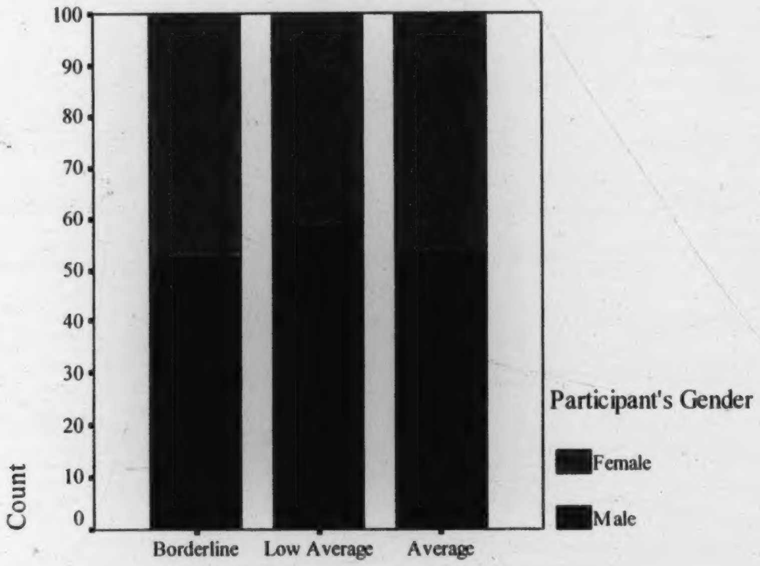
Figure 4.1

Student Classifications as a Percentage of Entire Sample (*N*=196)



Cases were selected within each IQ classification to provide a balance between males and females. Of the 196 cases included, 88 were female and 108 were males. In reviewing archived files, there appeared to be a preponderance of males completing psychoeducational assessments in the Education Clinic. Frequently more male than female cases were available but not all were included in the sample in an attempt to exert some control over the disproportionate size of gender-based groupings in subsequent data analysis. An attempt was made to control for age, gender and IQ classification to maintain numbers that did not differ excessively. For example, if males were going to exceed females within a classification, subsequent cases found were not selected as the search through archived files proceeded. Figure 4.2 illustrates the relative predominance of males in all three IQ classifications.

Figure 4.2 Comparison of IQ Classifications and Gender



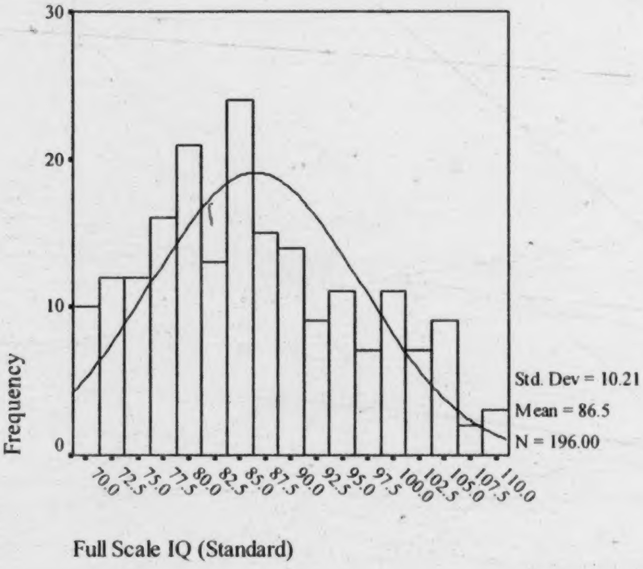
Student Classification According to Full Scale IQ

An attempt was made to obtain a representative range of IQs within each classification. Figure 4.3 provides an overview of the sample and the apparent positive skewing that appears to have resulted. In part, this could be explained by the fact that with both the Borderline and Low Average Classifications had a range of 10 IQ points while the

Average Classification covered a wider range with a full 20 IQ points. Therefore, the distribution of IQs in the sample favors the Average group given their overall numbers, contrary to what Figure 4.3 depicts. The possible influence of the Average IQ subpopulation on subsequent statistical procedures where IQ Classifications were used as either an independent variable or covariate resulted in employing procedures that controlled for unequal sample sizes.

Figure 4.3

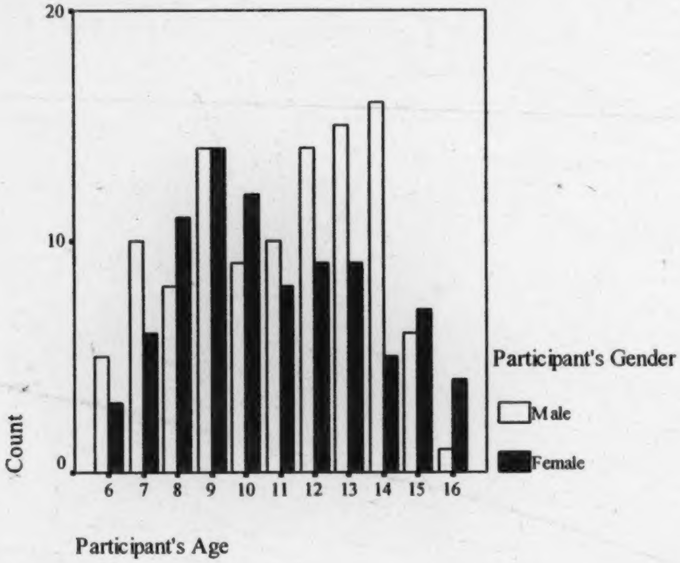
Histogram of WISC-III FSIQ range in Sample



In Figure 4.4 it can be observed that in six of the eleven age groupings the number of male participants exceeded females. More extreme differences in numbers were limited to just the 14 year old group where fewer females then males were obtained. Figure 4.5 provides an additional breakdown of the sample's participants according to grade and gender. Fewer numbers at both the upper and lower age ranges resulted in the majority of participants falling within the grade three to grade nine range.

Additionally, the individual's age was not considered when selecting cases for the research sample. Due to the fact that assessment referrals to the Education Clinic are a random process, the resulting archival assessment data makes no attempt to represent or misrepresent any particular age range. Should such a limitation be artificially imposed, then the total sample size would be reduced, effecting negatively both the power and significance of any subsequent statistical analysis.

Figure 4.4 Comparison of Gender and Age of Participants in Sample



An attempt was made to introduce more of a qualitative component to the sample's description in the form of information pertaining to grade retentions, enrolment in special programs and comorbid conditions. Of the 196 cases reviewed a total of 39 had repeated at least one grade (males – 24% versus females – 15%). An expected trend emerged where individuals with lower FSIQs tended to repeat a grade more frequently than those with higher FSIQs. The breakdown of individuals repeating at least one grade according to standard FSIQ classification was as follows: 31% of Borderline IQ, 21% of Low Average IQ and 10% of Average IQ.

In terms of special program description, eight general categories were established. It is acknowledged that there was little if any consistencies across the sample in terms of program information. However, both the contents of the archived file and the psychoeducational report were reviewed to obtain this information. Keeping in mind the possibility that the absence of reported programming may have been an error of omission, Table 4.1 summarizes the obtained descriptors. As expected, reported special programming was predominately in the Borderline and Low Average classifications. It is expected that the majority of individuals enrolled in modified, alternative or life skills programming would have some form of Individual Program Plan as mandated by the Special Education branch of Alberta Learning. Even with the low numbers of reported cases, Table 4.1 does show that more Borderline individuals had some formalized individualized education plan in place when compared to the Low Average and Average FSIQ groups. The changing conceptualizations of service delivery for students in need of some form of special education may be reflected in the low numbers of students attending a Resource Room within their regular school program. Inclusive programming for individuals with special educational needs together with the availability of modified and alternative curriculums may explain, in part, why so few students were reportedly attending a Resource Room.

Figure 4.5

Comparison of Gender and Grade of Participants in Sample

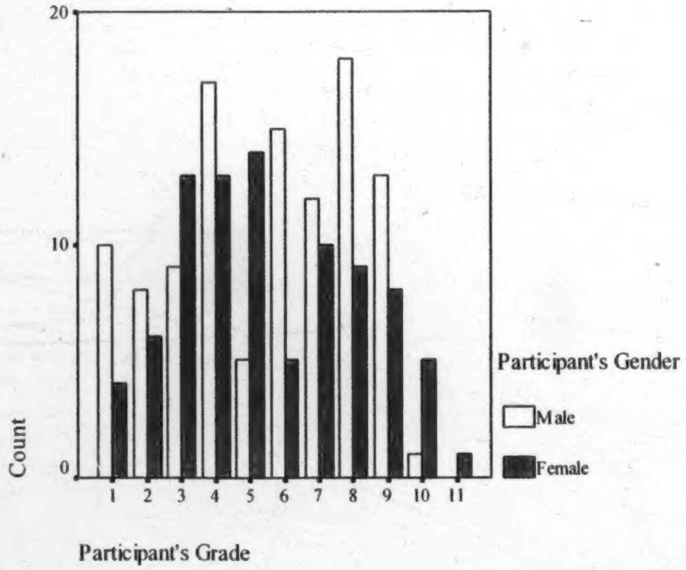


Table 4.1

Comparison of WISC-III FSIQ Classification and Educational Programming

	Borderline	Low Average	Average
No Special Program	17	22	50
Modified, Alternative, or Life Skills	16	19	5
IPP/IEP	11	3	3
Bilingual Program	2	4	6
Home Schooling	1	2	0
Resource Room	5	8	3
SNA	1	4	0
ESL Program	2	0	0

Note. SNA = Special Needs Assistant.

Table 4.2

Comparison of WISC-III FSIQ Classification and Comorbid Attentional, Behavioral, and Emotional Conditions

	Borderline	Low Average	Average	Total
None	14	12	17	43
MDD	6	0	0	6
LD	2	1	2	5
AD/HD	7	5	9	21
Attention Problems	6	16	15	37
Behavior/Attention	8	7	12	27
ODD	1	0	0	1
Behavior Problems	4	8	3	15
Depression/Anxiety	5	5	3	13
Multiple Diagnoses	2	6	5	13
Other	1	3	2	6

Note. MDD = Mild Developmental Delay; LD = Learning Disabled; AD/HD = Attention Deficit/Hyperactivity Disorder; ODD = Oppositional Defiant Disorder.

Table 4.2 summarizes additional qualitative information gleamed from individual assessment files. Roughly 23% of the sample had no reported emotional, behavioral or learning disorders. Similar to the previous table, a lack of reported information can not be accounted for given the archival nature of this study. Likely in many cases the reason for referral to the Education Clinic was to investigate the possible presence of these problems. The resulting psychoeducational assessment did prove, in some cases, to identify the presence of some of the conditions identified in the table.

Frequently the psychoeducational assessment reports would recommend further investigation through referral to other agencies and professionals rather than making definitive diagnostic conclusions. An additional problem encountered was the general lack of uniformity in the reporting of specific psychological problems or conditions. A case in point would be Disorders of Attention. Only 11% of the sample had a reported diagnosis of Attention Deficit/Hyperactivity Disorder (AD/HD) while in a remaining 20% of the files significant attentional problems were indicated. Together this represents at least 31% of the total sample where reported attentional problems, if not a DSM-IV diagnosed condition, were a significant concern impacting students' learning. If the group with combined behavioral and attentional problems is included with those individuals identified solely with attentional problems, then the resulting proportion of the entire sample grows to 45%. Similarly if individual cases with diagnosed Oppositional Defiant Disorder (ODD), Behavioral and Attentional problems as well as those with significant behavior problems are combined they represent roughly 23% of the entire sample. Typically within the Multiple Diagnoses group, one of the two or three of these problems related to behavioral or attentional concerns.

To summarize, it is clear that the majority of students included in the study had some form of learning, behavioral or emotional concerns. Of this estimated 77% of the sample, 22% fell into the Borderline FSIQ classification with the remaining 55% divided evenly between the Low Average and Average classifications.

Descriptive Results

Keppel & Zedeck (1989) underscore the importance of examining individual differences when they state, "variability is central to an understanding of data analysis" (p. 44). Initially, when examining results, attention turns to explicating similarities and differences with the end goal being the interpretation of the causes of observed variability.

Table 4.3

Summary of WISC-III Results for Sample and IQ-Based Classifications

	Full Sample (*N* = 196)		Average (*n* = 69)		Low Average (*n* = 68)		Borderline (*n* = 59)	
WISC-III	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
FSIQ	86.54	10.21	98.17	5.55	84.59	2.58	75.19	3.16
VIQ	87.59	9.70	98.48	5.81	85.13	3.26	77.69	4.05
PIQ	88.70	10.22	99.09	6.46	87.62	4.87	77.81	4.95
VCI	88.21	9.92	98.61	6.16	86.35	4.87	78.19	4.88
POI	90.26	10.63	100.04	7.60	89.63	6.30	79.54	5.91
FDI	87.65	12.82	96.71	11.86	84.04	11.03	81.19	9.59
PSI	92.66	14.44	101.00	11.21	91.06	14.76	84.76	12.82

Note. FSIQ = Full Scale Intelligence Quotient, VIQ = Verbal Intelligence Quotient, PIQ = Performance Intelligence Quotient, VCI = Verbal Comprehension Index; POI = Perceptual Organization Index; FDI = Freedom from Distractibility; PSI = Processing Speed Index.

Table 4.3 gives an overview of WISC-III results and demonstrates that expected trends were observed in relation to mean FSIQ, VIQ and PIQ for each of the three IQ-based classifications. Similarly, the Verbal Comprehension Index (VCI) and Perceptual Organization Index (POI) results did not show significant variability in results with means scores falling in expected ranges for each classification. One slight variation did occur with the FDI for the Borderline group where mean performance fell in the Low Average range. In this instance, less variability in individual scores is assumed given the standard deviation remained lower than the Average or Low Average groups as well as for the entire sample. Possibly a more significant result occurred with the Processing Speed Index (PSI) for the Low Average classification where the mean score fell in the Average. Here the standard deviation exceeded that of the entire sample and indicated significantly more variability in individual performance within the Low Average group than either the Average or Borderline groups for the PSI and remaining WISC-III results summarized in Table 4.3. The question remains, what is unique about the PSI that serves to differentiate individuals both within and between these IQ-based classifications? Subsequent inferential analysis will attempt to address this question.

Subtest Effect.

Sattler (2001) recommends that scaled scores ($M = 10$, $SD = 3$) for individual WISC-III subtests that range between 1 and 7 are to be classified as a weaknesses or below average while values from 8 to 12 are described as average. Scaled scores ranging from 13 to 19 are considered strengths or above average. According to these criteria the mean scaled scores for half of the WISC-III subtests in the sample (four Verbal Scale and two Performance Scale) fell into the below average range as illustrated in Table 4.4. However, if rounded to the nearest whole number only the Comprehension subtest mean (scaled score = 7) would remain classified as low average.

When IQ-based classifications are considered separately, the mean scaled score for each of the 12 subtests fell in the average range for the Average group and below average for the Borderline group as illustrated in Table 4.4. Results from the Below Average group were mixed with eight subtests (five Verbal Scale and three Performance Scale) falling in the below average range but with rounding to the nearest whole number, only Information, Arithmetic and Comprehension remained below average. The results of rounding scaled scores for the Borderline group would change the classification of three Performance Scale subtests (Digit Span, Picture Completion, Symbol Search) to average. To summarize, the mean performance on WISC-III subtests remained within expected levels for the Average and Borderline groups, whereas, considerably more variability in subtest results was observed in the Low Average group. For all IQ-based classifications Digit Span, Picture Completion, and Symbol Search results would fall into the average range if the rounding of group means is permitted. Typically scaled scores are not average but this was done as part of the exploratory analysis of the data in an attempt to understand further the nature of the individuals selected within each IQ classification.

Table 4.4

Scaled Score results of WISC-III Subtests for Sample and IQ-Based Classifications

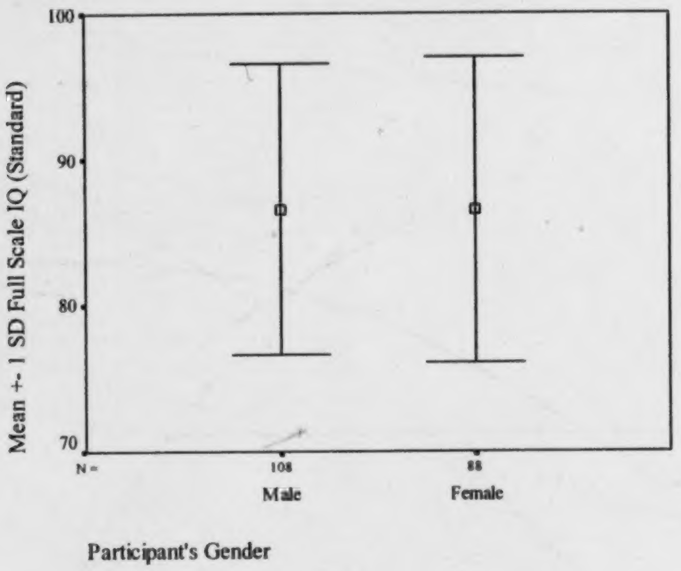
WISC-III Subtests	Full Sample (*N* = 196)		Average (*n* = 69)		Low Average (*n* = 68)		Borderline (*n* = 59)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Information	7.89	2.56	9.84	2.17	7.43	1.77	6.15	2.24
Similarities	8.61	2.43	10.45	1.97	8.32	1.78	6.80	2.05
Arithmetic	7.56	2.76	9.68	2.53	6.78	2.19	5.98	1.97
Vocabulary	7.94	2.49	9.93	1.85	7.75	1.72	5.85	2.07
Comprehension	7.42	2.57	9.01	2.13	7.25	2.13	5.76	2.42
Digit Span	8.20	2.53	9.01	2.69	7.85	2.46	7.64	2.18
Picture Completion	8.99	2.42	10.23	2.21	9.01	2.17	7.53	2.10
Coding	7.81	3.13	9.36	2.77	7.54	3.24	6.31	2.56
Picture Arrangement	8.72	3.00	10.28	2.56	8.84	2.69	6.78	2.74
Block Design	7.73	2.98	9.65	2.69	7.60	2.35	5.64	2.48
Object Assembly	8.15	2.61	9.77	2.33	7.97	2.11	6.47	2.30
Symbol Search	9.30	3.22	10.90	2.65	8.94	3.16	7.83	3.11

Gender.

As previously reported, the number of males exceeded females in all IQ classifications despite attempts to keep these differences small. The Borderline and Low Average groups were each comprised of 28 females while there was a total of 32 females with an Average FSIQ. Comparatively, the number of males was as follows: 31 Borderline, 40 Low Average, and 37 Average. As stated previously, eliminating cases with statistically significant Verbal and Performance Scale IQ differences served to reduce the number of available subjects for both genders. It also appears that relatively more males than females attend the Education Clinic for psychoeducational testing. Even with more males than females in the sample Figure 4.6 clearly indicates that the mean WISC-III Full Scale Intelligent Quotient (FSIQ) was quite similar. Additionally, some evidence of equality in the standard deviations of FSIQ scores between genders is also provided.

Figure 4.6

Average IQ of Males and Females in Sample



Age and Grade Levels.

Figures 4.7 and 4.8 illustrate the composition of each of the three IQ-based classifications. Given the random nature of referrals to the Education Clinic it is not possible to control the age or grade of individuals seeking psychoeducational assessments. Therefore when the Average, Low Average, and Borderline classifications are separated for comparative analysis it could be argued that developmental differences alone might be credited for resulting statistical differences given the apparent dissimilarities in group composition. To counteract this claim, apparent group similarities need to be examined closely.

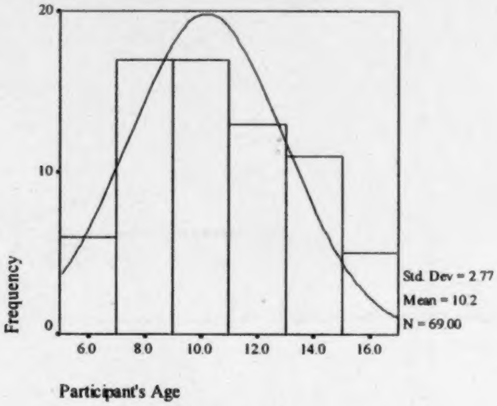
As Figure 4.7 pointed out, the average age for the Borderline and Low Average groups was almost identical (11.2 years and 11.3 years respectively). The age mean for the

Average group was only slightly lower at 10.2 years. Comparing the standard deviations of the age distributions of each IQ classification indicates that the greatest difference is only .33 years. In other words, the variability of scores around similar means differs by less than four months. Given the random nature of the sample, a difference of one year between the mean ages of IQ classifications will not likely adversely affect subsequent inferential statistics. The disproportionate number of retentions per IQ classification may prove to be more of a confounding variable.

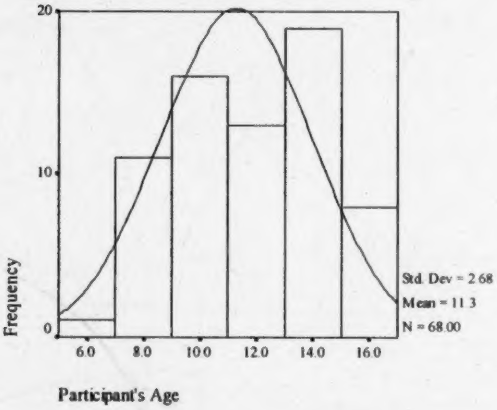
The average grade of participant for all classifications was within the grade five level although Figure 4.8 clearly indicates that the distribution with each IQ grouping varied somewhat. The mode for the Low Average group was grade eight while it was grade four and three for the Borderline and Average groups respectively. Younger and older ages are less well represented for all three classifications as depicted in Figure 4.7. The 'retention effect' becomes apparent when comparing the modes for age and grade across IQ classifications. Within the Average group the modes of eight years and grade three match expectations assuming most children start grade one at age six years. The distribution of the Borderline group deviates from this pattern with the mode for age (eleven years) exceeding the mode for grade (four). Although some of these individuals may have delayed entry into school, previous data indicated that 31 % reported at least one grade retention explaining why the modes for age and grade no longer match. With the multimodal age distribution (9 and 13 years) for Low Average individuals it is more difficult to clearly connect grade retentions at the 21% level with the mode of grade eight for this group.

Figure 4.7 Histograms of Age According to FSIQ Classification

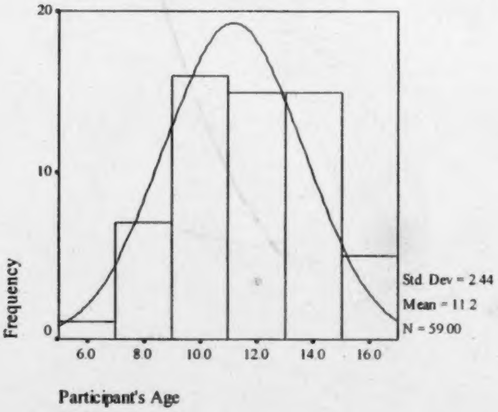
Age Distribution of Average IQ



Age Distribution of Low Average IQ

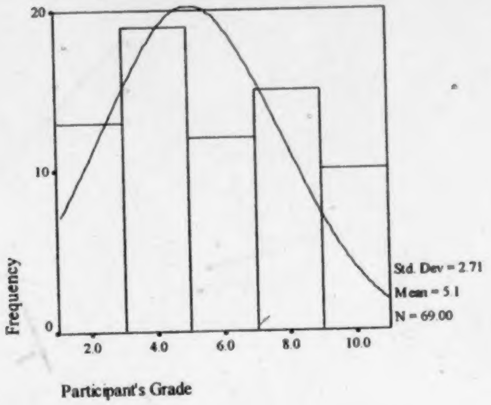


Age Distribution of Borderline IQ

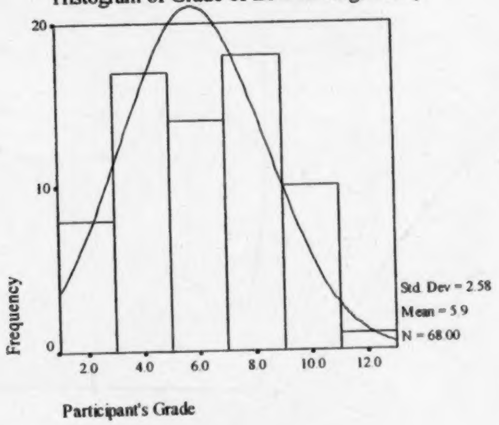


igure 4.8 Histograms of Grade According to FSIQ Classification

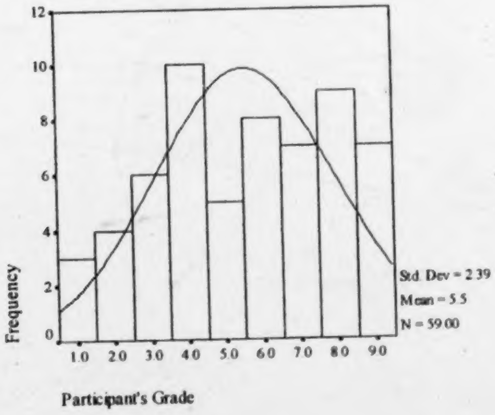
Histogram of the Grade of Average FSIQ



Histogram of Grade of Low Average FSIQ



Histogram of the Grade of Borderline FSIQ



Overall, the central clustering of age and grade ranges for the Average, Low Average and Borderline groups is evident. Additionally these distributions exhibited similar variability as reflected in the means and standard deviations of these independent variables.

Achievement Data Analysis

As part of the overall descriptive information for the research sample, Table 4.5 clearly illustrates that the WIAT was the most predominant measure of academic achievement utilized followed by the Canada Quiet as a distant second. Subsequent statistical analysis in the first research question took advantage of the relatively large subset formed with the WIAT to make comparative inferences when all the measures of achievement are grouped together.

Table 4.5

Achievement Test Administered According to IQ Classification

Achievement Test	Borderline	Low Average	Average	Total
WRAT	2	10	9	21
WJ-R	8	9	9	26
WIAT	31	37	35	103
Canada Quiet	17	10	16	43
WIAT-II	0	1	0	1
WJ-III	1	1	0	2
	59	68	69	196

Note: WRAT = Wide Range Achievement Test; WJ-R = Woodcock Johnson – Revised; WIAT = Wechsler Individual Achievement Test; WIAT-II = Wechsler Individual Achievement Test – Second Edition; WJ-III = Woodcock Johnson – Third Edition.

Given the nature of the archival information gathered it was not possible to obtain enough cases utilizing the same measure of academic achievement without compromising on more important variables such as elimination of cases with significant Verbal and Performance Scales IQs. Consequently, the combination of the five standardized achievement measures reported in Table 4.5 proceeded without sacrificing a significant reduction in sample size yet maintaining the construct validity of the data.

Reading versus Mathematics.

As Table 4.6 points out the mean Reading scores fell in the Average range for the sample while mean Mathematics scores fell in the Low Average range. A similar trend was also observed when achievement tests results were compared according to the three classifications. Table 4.7 illustrates that Reading achievement always exceeds Mathematics achievement for Average, Low Average and Borderline individuals.

Table 4.6

Achievement Test Means and Standard Deviations for Sample ($N = 196$)

Achievement Test	Word Recognition		Math Calculation		
	M	*SD*	*M*	*SD*	*n*
WRAT	92.48	13.33	86.38	11.91	21
WJ-R	100.00	13.43	86.38	15.03	26
WIAT	92.84	12.15	84.84	11.98	103
Canada Quiet	92.40	14.75	82.86	14.21	43
WIAT-II	82.00		79.00		1
WJ-III	100.50	.71	87.00	2.83	2
All Tests	93.68	13.15	84.77	12.80	196

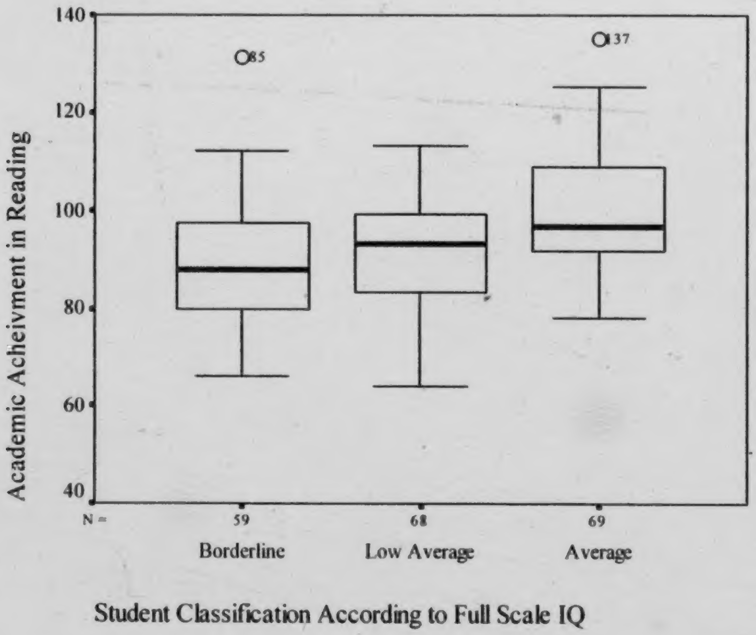
Table 4.7

Word Recognition and Mathematics Calculation Means and Standard Deviation for Full Sample (*N* = 196) and FSIQ Classifications

IQ Classification	Word Recognition		Math Calculation		
	M	*SD*	*M*	*SD*	*n*
Average	100.42	11.93	91.52	13.39	69
Low Average	91.29	11.97	81.75	11.33	68
Borderline	88.54	12.70	80.36	10.35	59
Full Sample	93.68	13.15	84.77	12.80	196

Figure 4.9

Boxplots of Reading Achievement According to IQ Classification



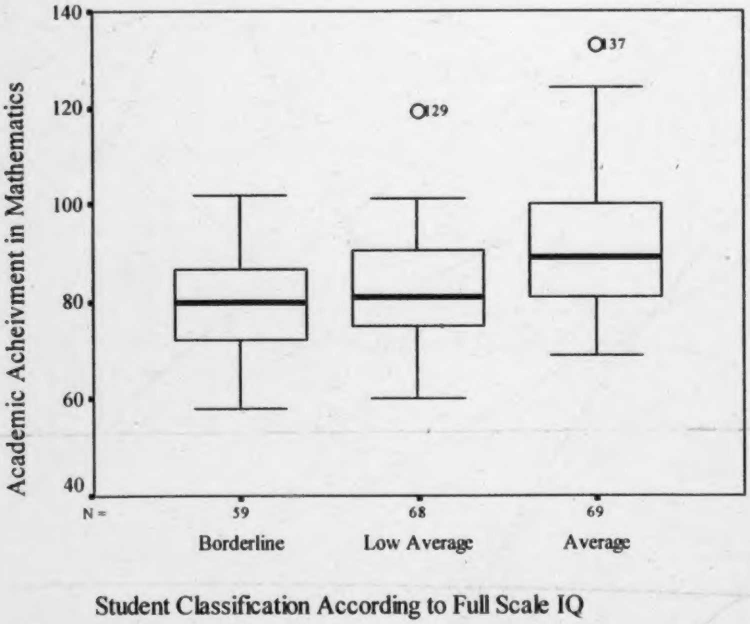
As part of the exploratory analysis of the data set Figures 4.9 and 4.10 provide some valuable insight into the distribution of mean Reading and Mathematics achievements scores, respectively. The fact that only three individuals out of a possible 196 would be considered outliers speaks to the homogeneity of variance within each of the three IQ classifications with respect to academic achievement levels. Additionally, by comparing the boxplots for the Borderline and Low Average groups in Figures 4.9 and 4.10 it is apparent that the range of scores is quite similar. For Mathematic achievement, Figure 4.10 also illustrates how close the median performance was for the Borderline and Low Average groups. Subsequent inferential statistics reported in Research Questions one, three and four provide consistent evidence that the Borderline and Low Average groups did not differ greatly with regard to mean achievement levels in reading and mathematics. In comparison, the Average group's achievement levels varied significantly from both the Borderline and Low Average groups' mean performance.

It is interesting to note that in figures 4.9 and 4.10 that the same individual, with an Average FSIQ, was the only outlier in both reading and mathematics achievement. Similarly, the two remaining outliers scored well above expected levels that might be predicted by their Low Average and Borderline FSIQs. Overall, the range of achievement scores within each IQ classification remained within expected limits and permit subsequent inferential statistics to proceed without the undue influence of a significant number of outliers and no extreme scores.

Combining a variety of achievement test scores is a confounding variable that could influence statistical analysis in a negative manner. However, from a practical perspective it is obvious that word reading tasks only differ on the targeted words used while the manner of presentation and student response remains quite similar. The calculation tasks assess knowledge of mathematical operations and compares individual results to widely accepted age and grade levels of mastery. Although research requires stringent controls to ensure validity and reliability, common sense must still prevail. It will be demonstrated that combining achievement test results is indeed legitimate and any observed variance in test scores within the same IQ classification is likely due to a host of variables, not just the test used to provide the scores.

Figure 4.10

Boxplots of Mathematics Achievement According to IQ Classification



Correlational Analysis

It is widely accepted that correlation coefficients are not necessarily indicative of causality (Witte & Witte, 2001). Furthermore, Sprinthall (1994) points out that significant correlations do not necessary communicate a "profound message" but rather these correlations are not likely the result of chance. Conclusions are further exacerbated when sample sizes increase resulting in smaller correlations attaining significance. All WISC-III subtest inter-correlations for the full sample ($N = 196$) are reported in Table 4.8. Despite the obvious preponderance of significant correlations it should be kept in mind that in terms of the actual clinical significance of these findings only a small portion of the variance is explained. In other words, even though low to moderate relationships between the majority of subtests has been demonstrated, the differential impact this could have on subsequent IQ-based classifications is yet to be determined.

Table 4.8

Inter-correlations Between WISC-III Subtests

Subtest	*1.*	*2.*	*3.*	*4.*	*5.*	*6.*	*7.*	*8.*	*9.*	*10.*	*11.*	*12.*
1 .Inf	1.00											
2.Sim	.37*	1.00										
3. Ari	.37*	.24*	1.00									
4.Voc	.35*	.49*	.22*	1.00								
5. Com	.20*	.24*	.20*	.30*	1.00							
6. DS	.10	.15*	.34*	.26*	.04	1.0						
7. PC	.31*	.27*	.26*	.22*	.29*	.09	1.0					
8. Cd	.11	.17*	.24*	.26*	.25*	.10*	-.01	1.00				
9. PA	.37*	.36*	.16*	.42*	.19*	.02	.10*	.07	10			
10.BD	.19*	.44*	.33*	.37*	.23*	.10*	.20*	.10	.20*	1.00		
11.OA	.28*	.41*	.34*	.38*	.20*	.15*	.14	.07	.11	.36*	1.00	
12. SS	.24*	.21*	.29*	.28*	.15*	.10	.14*	.45*	.12	.15*	.18*	1.00

Note: Inf = Information. Sim = Similarities. Ari = Arithmetic. Voc = Vocabulary. Com = Comprehension. DS = Digit Span. PC = Picture Completion. Cd = Coding. PA = Picture Arrangement. BD = Block Design. OA = Object Assembly. SS = Symbol Search.

*$p<.05$. **$p<.01$

Gender and Correlation.

In Table 4.9 significant positive correlations were found between the standard FSIQ, FSIQ (Symbol Search substitution), Intellectual Abilities (IA) and measures of word reading and mathematics calculation. This implies that as IQ increased proportionate increases in these achievement variables were also present in the full sample. Similar trends were also observed when female and male populations were considered separately. It is interesting to note that removal of the Arithmetic subtest in the IA was associated with a lowering of correlation coefficient levels in Mathematics achievement. This could add further proof that Arithmetic may prove to be a confounding variable, especially when attempting to utilize the IQ as a means of determining the presence of significant deficits in mathematics abilities.

Table 4.9

Correlations Between IQ, Achievement and Gender

IQ	Reading	Mathematics	Reading		Mathematics	
			F[b]	M[c]	F[b]	M[c]
FSIQ[a] (Standard)	.39**	.42**	.41**	.38**	.46**	.38**
FSIQ[a] (SS)	.39**	.40**	.40**	.39**	.43**	.38**
FSIQ[a] (IA)	.35**	.32**	.34**	.36**	.35**	.31**

Note. SS = Symbol Search substitution. IA = Intellectual Abilities (VCI and POI). F = Female. M = Male. Reading (Word Recognition subtest) and Mathematics (Calculation subtest) Achievement Measures: Wide Range Achievement Test; Woodcock Johnson – Revised; Wechsler Individual Achievement Test; Wechsler Individual Achievement Test – Second Edition; Woodcock Johnson – Third Edition; Canada Quiet.

[a]n = 196. [b]n = 88. [c]n = 108.

**p<.01

When the sample is subdivided into gender and IQ classifications the number of individuals per comparison is reduced as are the majority of significant correlations with achievement measures. Table 4.10 points out that it is mainly average males whose demonstrated significant associations with the Symbol Search substitution and Intellectual Abilities methods of calculating IQ for both reading and mathematics achievement. It should be kept in mind that the individuals comprising the three IQ groupings change somewhat as pointed out in greater detail in Research question 3. (c). As a result, comparisons between these groupings and subsequent achievement levels is restricted to more superficial analysis that merely points out that roughly the same general trends in correlation are maintained across all methods of IQ calculation.

Table 4.10

Correlations Between IQ Classifications, Achievement and Gender

IQ	Average		Low Average		Borderline	
	Female	Male	Female	Male	Female	Male
Standard FSIQ						
	(*n* = 32)	(*n* = 37)	(*n* = 28)	(*n* = 40)	(*n* = 28)	(*n* = 31)
Reading	.18	.30	-.27	-.28	.37	.07
Mathematics	.31	.41*	.08	-.26	.15	.02
FSIQ - SS						
	(*n* = 36)	(*n* = 39)	(*n* = 24)	(*n* = 39)	(*n* = 23)	(*n* = 22)
Reading	.33*	.51**	.07	.04	.32	.20
Mathematics	.31	.48**	-.10	-.20	.11	-.05
IA						
	(*n* = 34)	(*n* = 39)	(*n* = 27)	(*n* = 40)	(*n* = 24)	(*n* = 23)
Reading	.20	.41**	-.02	-.09	.32	-.24
Mathematics	.20	.40*	.08	-.03	.30	.15

Note. FSIQ = Full Scale Intelligence Quotient. SS = Symbol Search substitution. IA = Intellectual Abilities (VCI and POI).).

Reading (Word Recognition subtest) and Mathematics (Calculation subtest) Achievement Measures: WRAT; WJ-R; WIAT; WIAT-II; WJ-III, Canada Quiet.

*$p<.05$. **$p<.01$

However, one rather unexpected result did occur when the standard method of IQ calculation was subdivided into the 3 classifications with combined female and male groupings. Table 4.11 shows that a significant negative correlation with Reading was obtained for Low Average individuals when the standard method of IQ was utilized. For this select group, as their IQ increased within the 80 to 89 range proportionate decreases in levels of word reading ability were obtained. This is the exact opposite result when compared to the Average group where a significant positive correlation occurred. Results for the Borderline group were not significant. The Low Average groups' significant negative correlation with word reading ability does warrant further investigation into WISC-III cognitive profiles that may shed further insight into what might appear to be a spurious finding. The word reading task in question is devoid of contextual clues. Is it possible that Low Average readers may rely more heavily than their Average peers on contextual clues to recall individual words? Further discussion of these results will occur in the final chapter.

Table 4.11

Correlations Between IQ Classifications, and Achievement

IQ	*Average*		*Low Average*		*Borderline*	
	Re	Ma	Re	Ma	Re	Ma
FSIQ - Standard	($n = 69$)		($n = 68$)		($n = 59$)	
	.24*	.36**	-.26*	-.06	.23	.10
FSIQ - SS	($n = 75$)		($n = 63$)		($n = 45$)	
	.42**	.40**	.09	-.10	.25	.02
IA	($n = 73$)		($n = 67$)		($n = 47$)	
	.31**	.29*	-.08	.00	.11	.24

Note. FSIQ = Full Scale Intelligence Quotient. SS = Symbol Search substitution. IA = Intellectual Abilities (VCI and POI).). Re = Reading. Ma = Mathematics.

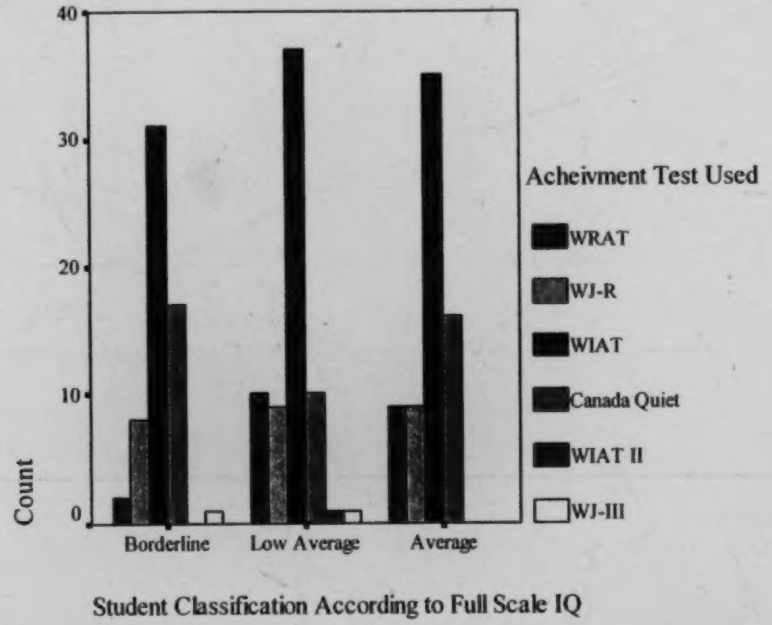
*$p<.05$. **$p<.01$

Correlation and IQ Calculation Method.

Of the five standardized achievement measures utilized in the study only the Wechsler Individual Achievement Test (WIAT) consistently demonstrated significant correlations between all three methods of IQ calculation (FSIQ, standard; FSIQ, symbol search substitution; IA or 8 subtest abbreviated IQ) and achievement levels in reading (word recognition task) and mathematics (calculation). The WIAT had been administered to a total of 103 individuals in the study. The remaining 93 subjects completed one of the five remaining standardized achievement tests: Wide Range Achievement Test ($n = 21$); Woodcock-Johnson Test of Achievement, Revised ($n = 26$); Canada Quiet ($n = 43$); Wechsler Individual Achievement Test, Second Edition ($n = 1$), and Woodcock-Johnson – Third Edition ($n = 2$). Figure 4.11 demonstrates that on the WJ-R, WIAT and Canada Quiet all three IQ classifications were well represented. The one exception being Borderline individuals completing the WRAT. Given the low overall numbers of participants completing the WRAT in the sample, it is unlikely that this deficit would have a significant impact on subsequent statistical analysis.

Figure 4.11

Achievement Test used According to IQ Classification



As expected, moderate correlations were observed between word recognition (WIAT) and IQ ($r = .514$ for FSIQ standard; $r = .519$ for FSIQ Symbol Search substitution; $r = .461$ for IA). Coefficients of determination or r^2 were 26%, 27% and 21% respectively. Therefore, approximately one quarter to one fifth of the variance in word recognition can be accounted for by the variability of IQ depending on the method employed to calculate this estimate.

A somewhat different trend emerged when correlation coefficients for mathematics achievement were examined. Low correlations were found between the WIAT Calculation subtest and all three methods of calculating IQ ($r = .396$ for FSIQ standard; $r = .376$ for FSIQ Symbol Search substitution; $r = .273$ for IA). However these small but definite relationships were significant at the 0.01 alpha level (2-tailed). The corresponding coefficients of determination were considerably lower than those reported for word recognition. The r^2 for mathematics achievement ranged from 16% (FSIQ standard) and 14% (FSIQ symbol search) to a low of only 7% when applying the abbreviated IQ formula (IA).

This pattern was not evident when mathematics achievement was measured by the Canada Quiet. Significant correlations demonstrated a moderate relationship between these mathematics achievement levels and IQ (*r* = .596 for FSIQ standard; *r* = .563 for FSIQ Symbol Search substitution; *r* = .527 for IA). The range for the resulting r^2 was slightly higher then what was reported for WIAT reading achievement and IQ. Coefficients of determination were 30% (FSIQ standard), 32% (FSIQ symbol search), and 28% (IA). Clearly, more of the variance in mathematics achievement on the Canada Quiet can accounted for by the variability in IQ than with the variance of Canada Quiet reading achievement levels and IQ. These latter correlation coefficients (Canada Quiet Word Recognition and IQ) were low and not significant (*r* = .273 for FSIQ standard; *r* = .296 for FSIQ Symbol Search substitution; *r* = .283 for IA). When comparing the scatter plots for the Canada Quiet results it is apparent that mathematic came closer to clustering around a linear plane than reading results that were more widely dispersed. Consistently lower levels of achievement in mathematics across IQ levels were more prevalent when compared to reading where both higher and lower scores tended to increase the spread between individuals' performance.

Inferential Statistics

This section starts with an overview of all four research questions before proceeding to the individual reporting of the statistical analysis for each one individually. An attempt was made to summarize some of the findings before proceeding to subsequent steps in the analysis of each null hypothesis although the chapter ends with a more detailed summary as well as a comparison of results from related questions.

Hypothesis One.

(a) There will be no difference in performance between individuals on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys.

(b) There will be no difference in performance between individuals on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline individuals grouped as Younger (ages 6, 7, 8, 9 years), Middle (ages 10, 11, 12) and Older (ages 13, 14, 15, 16).

Hypothesis Two.

Individual differences in factor indexes - Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), Freedom from Distractibility Index

(FDI), Processing Speed Index (FDI) – were examined to determine if Average, Low Average and Borderline students could be differentiated. The null hypothesis states that there is, in effect, no difference between IQ classifications when corresponding WISC-III factor scores are compared.

Hypothesis Three.

This portion of the study introduced a second method of calculating the Full Scale IQ (FSIQ) by substituting Symbol Search for Coding. The Null Hypothesis assumes there will be no difference in mean levels of performance on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys when the Symbol Search substitution (FSIQss) method is employed.

Hypothesis Four.

The fourth element of this study examined the effect of using an abbreviated IQ on subsequent classifications. Intellectual Abilities (IA) instead of the FSIQ was utilized to summarize WISC-III results. The Null Hypothesis assumes there will be no difference in mean levels of performance on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys when the IA, an abbreviated method of summarizing WISC-III results, is employed.

Testing Hypotheses One (a) and (b)

A 3 X 2 between subjects factorial design was employed to test the significance of IQ-based classification and gender on corresponding levels of academic achievement. Reading (Word Recognition) and Mathematics (Calculation) results will be considered separately to reduce the number of possible interaction effects. Similarly, a 3 X 3 between subjects factorial design was utilized to test the significance of IQ-based classification and age groupings on achievement levels.

It is recognized that the unbalanced factorial design utilized in the following two-way ANOVAs relies on unweighted mean to give equal weights to the different treatment groups in determining treatment effects even though a different number of observation may contribute to each treatment mean (Keppel & Zedeck, 1989). However, confidence in the following statistical findings can be gained by recognizing that the Type III sum of squares utilized by the data analysis program remains invariant to cell frequencies and "can be used with both balanced and unbalanced designs" (Field, 2000, p. 312).

Hypothesis 1. (a).

There will be no difference in performance between individuals on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys.

Descriptive statistics for reading achievement are presented in the appendix (Table A1) for each of the three IQ classifications according to the gender of the participant. The results from combining all the achievement tests and those obtained from just the Wechsler Individual Achievement Test (WIAT) are reported separately. Similar trends in word reading ability are observed with both the full sample and WIAT only groups when independent variables are considered. Statistically these similarities are well supported in a subsequent two-way ANOVA as reported in Table 4.12.

Table 4.12

Two-Way ANOVA (IQ Classification X Gender) for Word Reading Ability

	All Reading Tests Combined				WIAT only			
	(n = 196)				(n = 103)			
Source	*df*	*F*	*p*	*MSE*	*df*	*F*	*p*	*MSE*
Classification	2	16.70	.00**		2	13.70	.00**	
Gender	1	.110	.74		1	.47	.50	
C X G	2	1.14	.32		2	.15	.87	
within group	190			148.84	97			117.18

Note. WIAT = Wechsler Individual Achievement Test. Classification = Borderline, Low Average, Average. C X G = Interaction of IQ Classification and Gender.

** $p < .01$

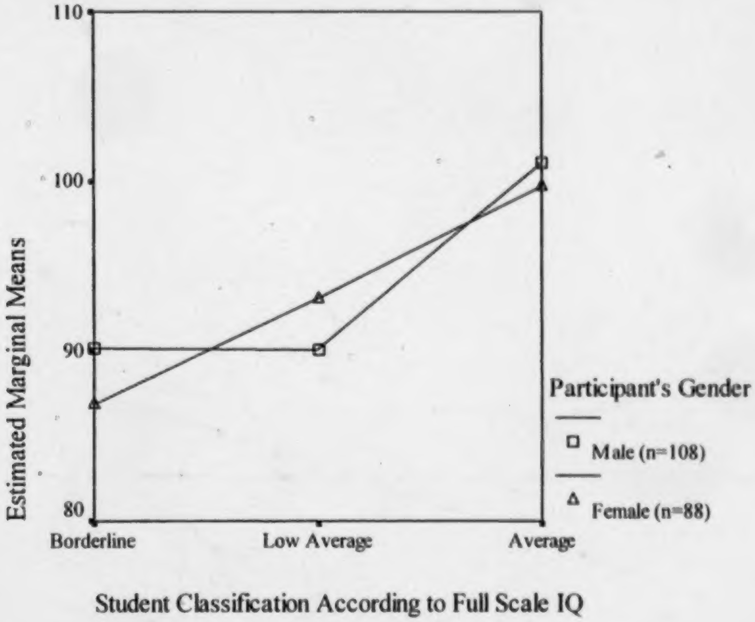
Results from the 3 X 2 between subjects factorial ANOVA presented in Table 4.12 would lead to rejection of the null hypothesis for research question 1. (a) that pertains to IQ classification but acceptance of the null hypothesis for gender. Overall, significant differences in average word reading ability does exist between Borderline, Low Average and Average individuals ($F(2,196) = 16.70, p = .01$) but males and females average reading abilities were not significantly different ($F(2,196) = 16.70, p = .01$). It does not appear that combining five different measure of word reading had a significant impact on these results given that exactly the same trend was evidenced in the WIAT only subsample. Levene's test of equality of error variances was not significant which lends more support to the validity of these results.

All forms of Post Hoc analysis (Tukey HSD, Scheffe, Bonferroni, Games-Howell) produced the same results. Specifically, mean word reading ability was significantly different when the Average and Low Average groups were compared as well as between the Average and Borderline groups. However, this was not the case when Borderline and Low Average groups were compared and it was found that mean performance on word reading tasks did not differ significantly. Figure 4.12 clearly illustrates these results. Reading achievement results appears to depend on IQ classification but not gender, with no evidence of a significant interaction between these independent variables as illustrated in Table 4.12.

With regard to the second part of the same research question that considered the participant's performance in mathematics rather than reading, the descriptive statistics are presented in the appendix (Table. A2). In comparison to Table A1, all levels of Mathematics performance were lower than reading levels. It could be argued that performing mathematical operations is cognitively more challenging than a simple word recognition task. Perhaps more importantly the same statistical results occurred despite these obvious differences in task demands.

Figure 4.12

Estimated Marginal Means of Word Reading by IQ Classification and Gender



Testing for the main effect of IQ classification and mathematics achievement levels produced statistically significant results for the full sample, *F*(2,196) = 17.25, *p* = .00 as well as the WIAT only subsample, *F*(2,97) = 12.54, *p* = .00 as illustrated in Table 4.13. Levene's test was not significant in either case confirming that error variance was not excessive across the three IQ groupings. As with the reading achievement levels, post hoc multiple comparisons found no difference in mean mathematics achievement levels between the Borderline and Low Average groups. Instead, the significant differences between the Average group and the remaining two classifications were likely responsible for producing the overall main effect of IQ and mathematics achievement. This trend is graphically illustrated in Figure 4.13.

To briefly summarize the results of research question 1 (a), the null hypothesis would not be rejected for both reading and mathematics achievement when gender was utilized as an independent variable. That is to say, mean levels of academic achievement did not differ significantly when the girls' and boys' results were compared. There was no interaction between gender and IQ classification which gives more weight to the significant findings for the later. When all achievement tests were grouped or when just the WIAT was considered separately, some statistically differences in performance levels between IQ classifications were confirmed. Specifically, those with Average IQ were found to have significantly higher word reading and calculation abilities then either their Low Average or Borderline peers. Although some variation in levels of performance was observed when the Low Average and Borderline groups were compared, these differences were not found to be statistically significant.

An obvious limitation in these findings relates to differences in samples sizes for both gender and IQ groupings. The robust nature of the ANOVA calculations in these circumstances is reinforced by confirmations of variance homogeneity within all levels of analysis. Given the random nature of client's attendance at the Education Clinic, the resulting archival data gathered from these files was unable to control for balanced numbers of participants as is expected of pure experimental research. In addition, the careful selection of files eliminated numerous cases but served to prevent the introduction of more serious and potentially more damaging confounding variables into the study.

Hypothesis 1. (b).

There will be no difference in performance between individuals on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline individuals grouped as Younger (ages 6, 7, 8, 9 years), Middle (ages 10, 11, 12) and Older (ages 13, 14, 15, 16).

The same procedures were followed to investigate the effect of age on IQ classification and achievement levels in the second portion of the first research question. The descriptive statistics presented in the appendix (Table A3) compare word recognition levels across the three age groupings according to the three IQ classifications. The obvious variance differences in reading scores as reflected in the standard deviation columns of this table may help explain the significant Levene test for homogeneity. The wide range in cell sample sizes from a low of *n* = 16 to a high of *n* = 33 within the full sample not only exacerbates observed differences in reading levels but adds to the instability of statistical findings. For these reasons the results reported in Table 4.14 should be considered cautiously given these aforementioned limitations.

With regard to age groupings and reading achievement, the null hypothesis would be accepted indicating that observed differences in scores were not statistically significant. Conversely, the alternate hypotheses is accepted when just IQ classification is considered, $F(2,196) = 16.25, p = .01$ for the full sample and just the WIAT, $F(2,94) = 15.73, p = .01$. The interaction between age and IQ groupings was not statistically significant. Results of post hoc comparisons were the same as with part (a) where differences between the Average and Low Average or Average and Borderline classifications were found to be significant. Figure 4.14 illustrates the unexpected trend where the middle group's mean reading abilities is lower than the younger groups' results across all three IQ classifications. Sample size differences introduce a limitation that makes it difficult to interpret these findings. A significant Levene's test of the equality of error variances also calls into question the reliability of these statistical comparisons.

Table 4.14

Two-Way ANOVA (IQ Classification X Age Group) for Word Reading Ability

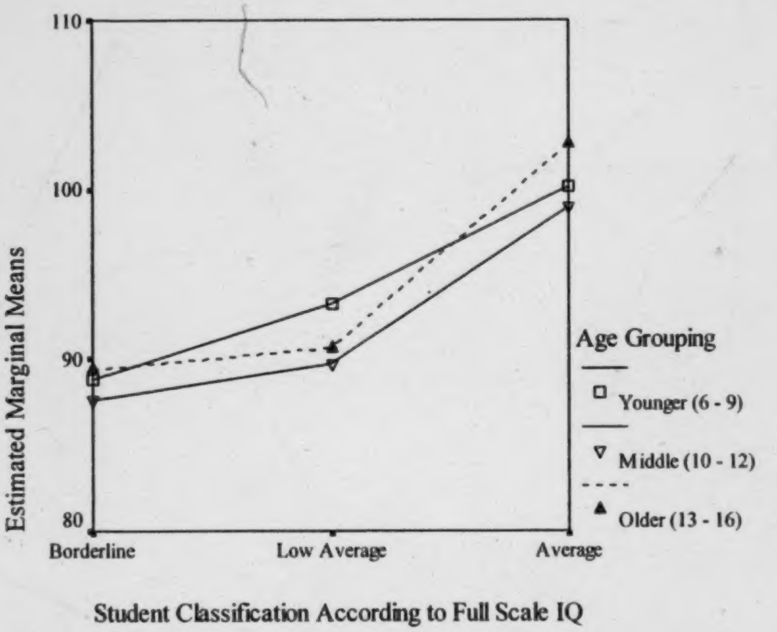
	All Reading Tests Combined (*n* = 196)				WIAT only (*n* = 103)			
Source	*df*	*F*	*p*	*MSE*	*df*	*F*	*p*	*MSE*
Classification	2	16.25	.00**		2	15.73	.00**	
Age (group)	2	.57	.57		2	2.56	.08	
C X A	4	.26	.90		4	.77	.55	
within group	187			151.47	94			112.66

Note. WIAT = Wechsler Individual Achievement Test. Classification = Borderline, Low Average, Average. C X A = Interaction of IQ Classification (Borderline, Low Average, Average) and Age Group (younger, middle, older).

** $p < .01$

Figure 4.14

Estimated Marginal Means of Reading by IQ Classification and Age Group



Age groupings had a different effect then gender groupings when Mathematics achievement levels were compared across IQ classifications. As depicted in the appendix (Table A4), the descriptive statistics point to an inverse relationship between the independent variables. In other words, as the age of the participant increased, their corresponding mathematics abilities decreased. This was consistent across all three IQ classifications. This lead to rejection of the null hypothesis, indicating that a main effect for age grouping was statistically significant, $F(2,187) = 9.27, p = .01$ for the full sample as well as with just the WIAT results, $F(2,94) = 7.61, p = .01$. Table 4.15 confirms that the interaction of age grouping and IQ classification was not significant. Unlike the previously reported findings for word reading ability, Levene's test was not significant for mathematics calculation. Therefore the illustrated trends depicted in Figure 4.15 could be interpreted with more confidence that observed differences in mean mathematics achievement levels is not due, in part, to unequal variances in scores when age groupings and IQ classifications are compared.

Table 4.15

Two-Way ANOVA (IQ Classification X Age Group) for Mathematics Calculation

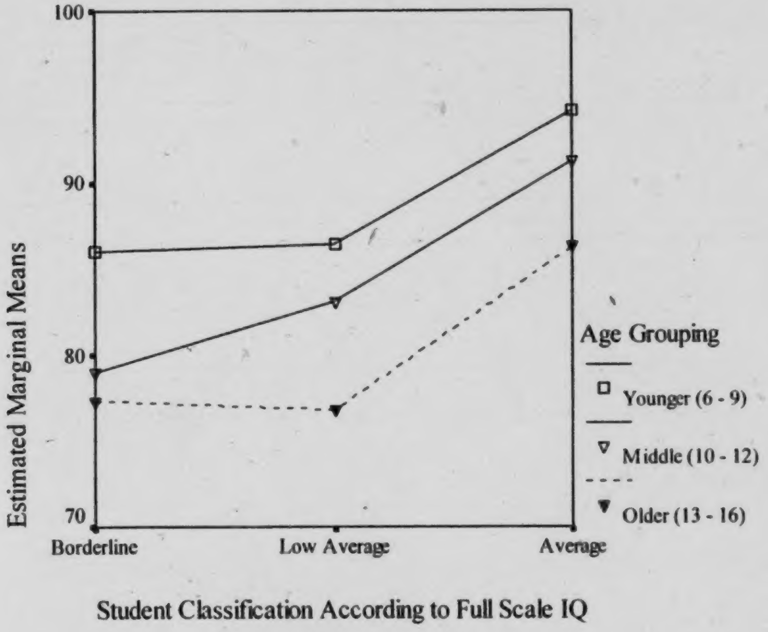
	All Math Tests Combined (*n* = 196)				WIAT only (*n* = 103)			
Source	*df*	*F*	*p*	*MSE*	*df*	*F*	*p*	*MSE*
Classification	2	13.40	.00**		2	10.95	.00**	
Age (group)	2	9.27	.00**		2	7.61	.00**	
C X A	4	.32	.86		4	1.47	.22	
within group	187			151.47	94			100.65

Note. WIAT = Wechsler Individual Achievement Test. Classification = Borderline, Low Average, Average. C X A = Interaction of IQ Classification (Borderline, Low Average, Average) and Age Group (younger, middle, older).

** *p* < *.01*

Figure 4.15

Estimated Marginal Means of Math Calculation by IQ Classification and Age Group



To summarize the results of research question 1 (b), the null hypothesis would be accepted for reading achievement but rejected for mathematics achievement when age groupings were utilized as an independent variable. As with research question 1 (a), differences in numbers of individuals included within comparison groups likely contributed to the significant level of error variance observed when word reading ability was the dependent variable. More homogeneity in variance was present among age groupings when mathematics ability was considered.

The inverse relationship between age grouping and level of calculation ability proved to be a statistically significant finding for this portion of the research question. All post hoc comparisons were significant suggesting that this main effect for age groupings was consistent throughout all possible combinations of the younger, middle and older age groupings. It was confirmed that no significant interaction occurred with either the full sample or WIAT only subsample. As a result, it appears that mathematics achievement levels did not depend on a combination of age and IQ classification. Instead, it was obvious that Average individuals achievement levels in mathematics were consistently greater than their Low Average and Borderline peers with older individuals tending to score significantly lower then individuals from either the middle or younger groups.

Testing Hypothesis Two

> *Individual differences in factor indexes - Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), Freedom from Distractibility Index (FDI), Processing Speed Index (FDI) – were examined to determine if Average, Low Average and Borderline students could be differentiated. The null hypothesis states that there is, in effect, no difference between IQ classifications when corresponding WISC-III factor scores are compared.*

Intuitively, the alternative hypothesis would appear to be more logical and One-Way ANOVAs did prove this to be the case when the full sample was considered. Levene's test for the homogeneity of variance was significant for the POI but not the remaining three factor indexes. Violation of this assumption will call into question subsequent ANOVA results. The fact that more variation in subtest results would have occurred within the POI to produce this result warrants closer examination.

It may be assumed that less variation in WISC-III Verbal Scale results is reflected in the VCI since Levene's test was not significant for this factor. As a result it appears that more variability in performance was evident within and across IQ classifications for nonverbal than verbal processing tasks.

However, the mean levels of performance on the four subtest Verbal and Performance Indices stayed within expected classification boundaries for the Borderline, Low Average and Average groups as confirmed by Table A5 in the appendix. This was not the case for the two-subtest PSI and FDI where IQ classification and mean performance did not always align. This, perhaps, is an unfair if not unreliable comparison given the composition of four versus two subtest indices, yet from a School Psychology perspective psychoeducational reports typically report all Factor Indices and highlight significant strengths and weaknesses that may emerge (Sattler, 2001).

To account for unequal samples sizes between the Borderline, Low Average and Average groupings, both the unweighted and weighted means were utilized in the ANOVA. As Table 4.16 illustrates all *F* tests remained significant suggesting that the Null hypothesis should be rejected. This indicates that not all Factors Indexes were consistent with the global FSIQ classification from which they were derived.

Table 4.16

Univariate *F* Tests for IQ Classification and WISC-III Factor Scores

Source	*df*	*F*	*p*	*MSE*
VCI	2	236.92**	.000	28.75
POI	2	150.33**	.000	44.60
PSI	2	26.02**	.000	165.97
FDI	2	37.73**	.000	119.35
Within Groups	193			
Total	195			

Note. VCI = Verbal Comprehension Index. POI = Perceptual Organization Index. PSI = Processing Speed Index. FDI = Freedom from Distractibility Index.

** $p < .01$

In addition to the Tukey's Honestly Significant Difference (HSD) post hoc analysis the Games-Howell procedure was utilized due to the fact that it accounts for unequal sample sizes. All resulting comparisons were significant for each of these procedures across the VCI, POI, and PSI factors. In other words, the differences in mean performance between the Borderline, Low Average, and Average groups for VCI, POI and PSI was statistically significant at the $p > .01$ level. The implication is that the VCI, POI, and PSI will consistently distinguish individuals from these three IQ classifications. The exception was the FDI where only the Borderline and Average groups differed significantly.

The Tukey HSD, Games-Howell procedures, Hochberg's GT2 and Gabriel's pairwise test procedures were utilized to account for the unequal sample sizes. As a result, all forms of post hoc analysis suggested that the Null Hypothesis was true when the mean FDIs for the Low Average and Borderline groups were compared. As illustrated in the appendix (Table A6), the mean FDIs for these two classifications fell in the Low Average range with less than three scaled score points separating them. The FDI is comprised of two Verbal Scale WISC-III subtests, Arithmetic and Digit Span, with the latter being a supplemental subtest, not included in the calculation of the FSIQ.

The FDI was the only factor that did not serve to differentiate the Borderline and Low Average groups. While on the surface, the Null Hypothesis was rejected for the full sample ($F(2, 193) = 51.16, p > .01$) for weighted means, post hoc comparisons revealed that the Null Hypothesis would be accepted when comparing mean FDI results for these groups. Unlike the PSI, no gender effect was found for the FDI. In other words, the Null Hypothesis continued to be accepted even when comparisons between Borderline and Low Average classifications were completed according to gender groupings. Table A6, in the appendix, compares FDI and subtest results. For both Borderline and Low Average groups Arithmetic is about one scaled score point less than Digit Span in comparison to the Average group where mean performance is more equal.

To determine the significance of gender and mean WISC-III factor score results a One-Way ANOVA was completed to compare Borderline, Low Average and Average groups. Similar to the full sample, expected trends emerged for both females and males where mean factor scores differed significantly when the three IQ-based classifications were compared. Unlike the full sample, Levene's test for the homogeneity of variance was not significant for any of the mean factor scores. The effect of separating females' and males' results on the POI eliminated the previously reported significance for Levene's test.

In completing post hoc comparisons unequal sample sizes were accounted for by Hochberg's GT2 and Gabriel' pairwise test procedures. The same levels of significance (p >.01) as observed in the full sample were obtained with all combinations of gender and classification for the VCI and POI factors scores. Similar to the full sample, the FDI was not significant for females and males when the Borderline and Low Average groups were compared. Unlike the full sample, gender had an effect of identifying differential performance on the PSI. It appears that Borderline and Low Average males PSI results were more similar than Borderline and Low Average females where significant differences on all post hoc comparisons were observed. Table A7, in the appendix, illustrates that the mean level of performance on the Symbol Search task was similar for the Borderline and Low Average groups. Keeping in mind that Symbol Search, in the full sample, more closely approximates a normal distribution as compared to Coding, subsequent conclusions drawn from analysis of the effect of IQ-based classification on these distributions can stand on more firm statistical ground.

Testing Hypothesis Three.

> *This portion of the study introduced a second method of calculating the Full Scale IQ (FSIQ) by substituting Symbol Search for Coding. The Null Hypothesis assumes there will be no difference in mean levels of performance on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys when the Symbol Search substitution (FSIQss) method is employed.*

As part of the third general research question, the determination of the Full Scale IQ (FSIQ) was changed slightly. Following Kaufman's (1994) recommendation that substituting Symbol Search for Coding become a routine practice was facilitated by the use of Canadian norms (Saklofske, Hildebrand, Reynolds, & Wilson, 1998). The null hypothesis states that this procedure would have no significant effect on the resulting FSIQ with Symbol Search replacement (FSIQss) as compared to the standard FSIQ. Subsequently the classification of individuals according to their FSIQss would not differ significantly from standard FSIQ classifications. Alternatively if classification changes did occur, then the nature of the Symbol Search and Coding subtests would need to be examined closely to tease out implications related to intelligence and information processing.

Table 4.17

Effect of Symbol Search Replacement FSIQ on IQ Classifications

	Change in Classification after SS Replacement			
	No Change	Increase	Decrease	Total
Borderline	42	8	9	59
Low Average	54	11	3	68
Average	64	4	1	69
Total	160	23	13	196

Note. SS = Symbol Search. FSIQ = Full Scale Intelligence Quotient.

Figure 4.16

Significant Differences Between Symbol Search and Coding

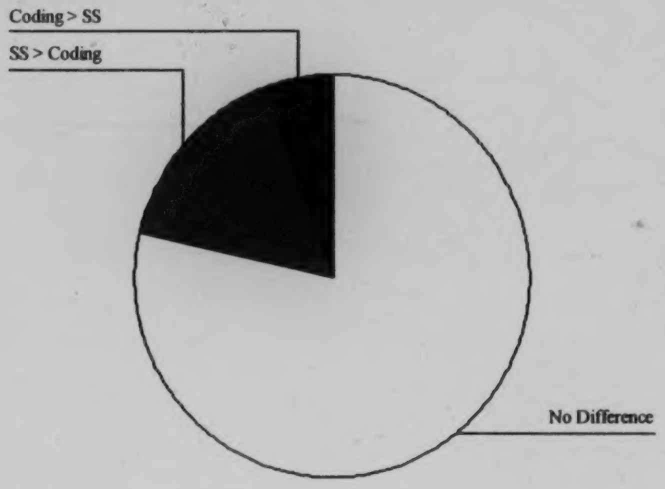
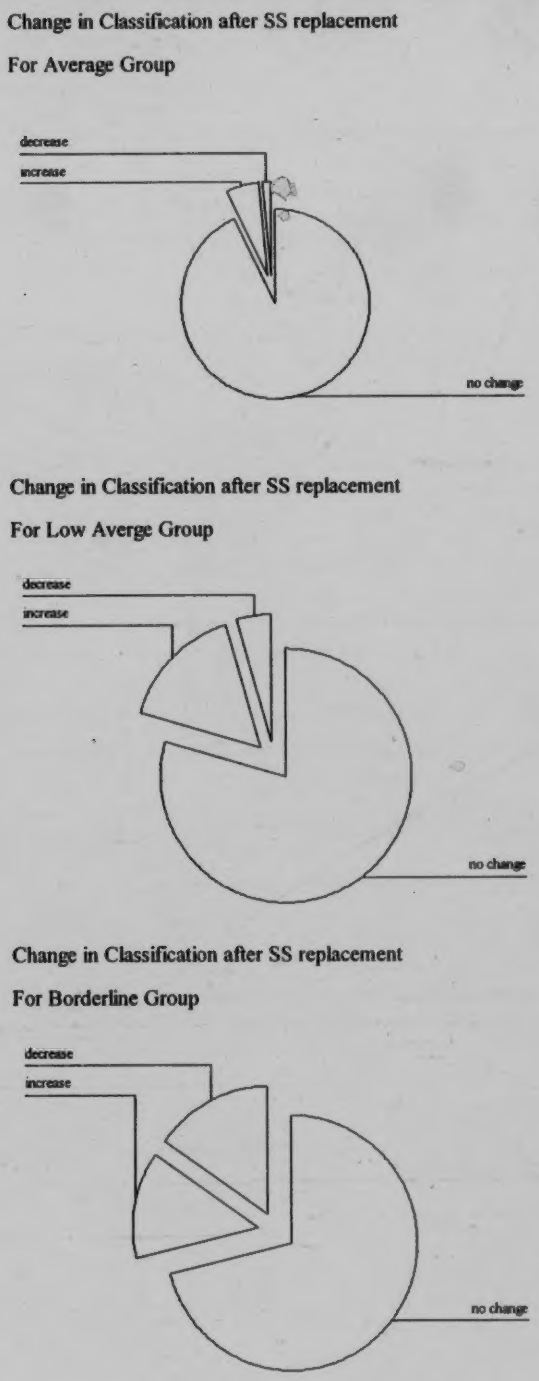


Table 4.17 represents the effects of performing the Symbol Search substitution with the number of individual FSIQs effected showing some variation when IQ-based classifications are considered. This classification would either increase or decrease for 28% of Borderline, 20% of Low Average and only 7% of Average individuals in the sample. With substantially more increases than decreases evident within the Low Average group it is apparent that Symbol Search was more of a strength than Coding for 16% of this group. This same pattern was not repeated with either the Borderline or Average groups.

Prifitera, Weiss & Saklofske (1998) point out that a difference or five points or more will invalidate resulting Processing Speed Index (PSI) calculations. Figure 4.16 illustrates that with 21% of the sample the PSI should not be considered valid for this reason. Accordingly, if Symbol Search is substituted in the calculation of the FSIQ, it will increase resulting scores in 15% of the sample. When cases with significant Symbol Search and Coding differences were excluded the resulting mean PSIs did not differ greatly from the PSIs for the full sample. Subsequently, the classification of mean PSIs remained unchanged: Low Average for the Borderline group, Average for the Low Average group and Average for the Average group.

Figure 4.17 Changes According to Classification Resulting from Symbol Search Replacement



Change in Classification after SS replacement

For Average Group

Change in Classification after SS replacement

For Low Averge Group

Change in Classification after SS replacement

For Borderline Group

Determining an effective way to analyze this research question was problematic initially. Figure 4.17 provides visual inspection of the classification changes following the Symbol Search substitution method of determining the FSIQ. Initially the sample selection for the study started with three IQ classifications and Table 4.17 clearly shows the Average group had the fewest changes in classification with the Borderline and Low Average groups experiencing considerable more changes in classification (increases or decreases).

The addition of the High Average and Extremely Low classifications resulted following the employment of these alternate methods of summarizing WISC-III performance. As Table 4.18 points out the number of individuals reclassified as High Average or Extremely Low were not great, with only four and nine changes observed respectively. The 11 Low Average individuals reclassified as Average represents the most changes overall. Consequently, attention returns to the three main classifications of Average, Low Average and Borderline, where a general shifting in individuals falling within these classifications occurred depending on FSIQ methodology. Arguably, one premise of the study continues to be reinforced when it is recognized that IQ-based classifications may not accurately reflect an individual 'true' level of intelligence. Rather, a quasi-sophisticated method of transforming individual scores to normative scores is provided by the test developers. One's estimated level of intellectual functioning is dependant on the manner in which the raw scores are transformed to standard scores as much as the individual items used to accentuate differences in presumed ability.

The main purpose for including the Symbol Search substitution and Intellectual Ability methods in the study was to highlight the malleability of the FSIQ as well as the corresponding IQ-based classifications. With so much emphasis placed on IQ-cutoff scores in school systems, psychoeducational reports would be remiss if they failed to clearly report these classifications, especially when needed government funding may be at stake. In an attempt to mimic this same hard line approach, a decision was made to treat the revised classifications of Average, Low Average, and Borderline in the same manner as which a government auditor might as they scrupulously review an individual's test results and corresponding recommendations for special needs coding. In other words, a standard score is a standard score, no matter how it was obtained.

Table 4.18

Comparison of IQ Classifications between Standard FSIQ and FSIQss Methods

	Classification According to Standard FSIQ			
FSIQss Classification	Borderline	Low Average	Average	Total
Extremely Low	9	0	0	9
Borderline	42	3	0	45
Low Average	8	54	1	63
Average	0	11	64	75
High Average	0	0	4	4
Total	59	68	69	196

Note. FSIQ = Full Scale Intelligence Quotient. FSIQss = Full Scale Intelligence Quotient Symbol Search Substitution.

As Table 4.18 illustrates, only a small portion of the individuals changed classifications as a result of utilizing the Symbol Search subtest rather than Coding to calculate the FSIQ. Given the majority of the individuals included in the three classifications (Average, Low Average, Borderline) are the same as when the standard FSIQ is utilized, it would be of research interest to determine if these FSIQss classifications changes might lead to changes in results found in the first research question.

The same 3 X 2 between subjects factorial design employed in the first research question was utilized to test the significance of IQ-based classification based on Symbol Search substitution with gender groupings on corresponding levels of academic achievement (Word Recognition) and (Calculation).

Results from the first research question were two-fold. The null hypothesis was rejected for IQ classifications with significant differences found between the Average

Table 4.13

Two-Way ANOVA (IQ Classification X Gender) for Mathematics Calculation

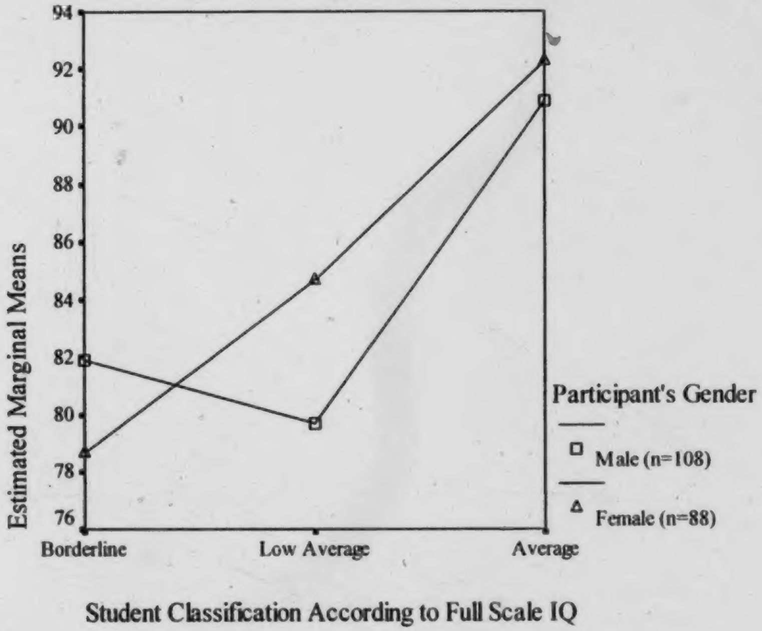
	All Math Tests Combined (*n* = 196)				WIAT only (*n* = 103)			
Source	*df*	*F*	*p*	*MSE*	*df*	*F*	*p*	*MSE*
Classification	2	17.25	.00**		2	12.54	.000**	
Gender	1	.39	.531		1	3.91	.051	
C X G	2	1.87	.157		2	.61	.544	
within group	190			139.09	97			114.33

Note. WIAT = Wechsler Individual Achievement Test. Classification = Borderline, Low Average, Average. C X G = Interaction of IQ Classification and Gender.

** $p < .01$

Figure 4.13

Estimated Marginal Means of Math Calculation by IQ Classification and Gender



The third research question produced similar results as the first research question. The 3 X 2 between subjects factorial ANOVA presented in Table 4.20 would lead to rejection of the null hypothesis that pertains to IQ classification but acceptance of the null hypothesis for gender. Levene's test of equality of error variance was not significant, With the homogeneity of variance assumption maintained, further confidence in the interpretation of the following results is permitted despite obvious differences in group numbers for classification and gender as well as a slightly reduced sample size (*n* = 183). To account for some of these differences a variety of post hoc comparisons were employed (Tukey HSD, Scheffe, Gabriel, Hochberg, and Games-Howell) with a consensus obtained that reproduces the findings of the first research question.

Overall, significant differences in mean word reading ability does exist between Borderline, Low Average and Average individuals (*F*(2,183) = 6.42, *p* <.01) but males and females mean reading abilities were not significantly different (*F*(2,183) = .05, *p* <.82). When the Low Average and Borderline groups were compared, Reading achievement levels were not significantly different while the Average group's superior abilities reached statistically significance levels.

Table 4.20

Comparing Two-Way ANOVAs (IQ Classification X Gender) for Word Reading Utilizing Standard FSIQ and FSIQss

	Standard Full Scale IQ (*N* = 196)				Symbol Search Substitution FSIQ (*n* = 183)			
Source	*df*	*F*	*p*	*MSE*	*df*	*F*	*p*	*MSE*
Classification	2	17.25	.00**		2	6.42	.00**	
Gender	1	.39	.531		1	.05	.82	
C X G	2	1.87	.157		2	1.09	.34	
within group	190			139.09	177			160.92

Note. WIAT = Wechsler Individual Achievement Test. Classification = Borderline, Low Average, Average. C X G = Interaction of IQ Classification and Gender.

FSIQss = Full Scale Intelligence Quotient with Symbol Search Substitution.

** *p < .01*

Results from the second part of the third research question that considers the participant's performance in mathematics, mirrors that of the first research question. The descriptive statistics are presented in Table. 4.21. In comparison to Table 4.19, all levels of Mathematics performance were lower than reading levels. As Table 4.22 illustrates, substituting Symbol Search for Coding in determining the FSIQ did not change the resulting statistical findings where gender did not serve to differentiate individuals while significant differences in Mathematics achievement levels were more obvious when IQ classifications were considered. Consistent with all previous findings, the differences were found to lie between the Average group and the remaining classifications. Although mean differences in Mathematics calculation abilities were evident between the Borderline and Low Average groups they were not statistically significant.

Table 4.21

Comparison of Mathematics Achievement Standard Score Means and Standard Deviations for Standard IQ and Symbol Search Substitution Classifications According to Gender

	Standard Full Scale IQ					
	Borderline		*Low Average*		*Average*	
Math (all tests)	*M*	*SD*	*M*	*SD*	*M*	*SD*
Female	78.68[a]	10.11	84.68[b]	12.81	92.28[c]	14.19
Male	81.87[d]	10.50	79.70[e]	9.83	90.86[f]	12.83
	Symbol Search Substitution FSIQ					
Math (all tests)						
Female	78.57[g]	8.99	85.92[h]	13.10	91.11[i]	14.19
Male	80.32[j]	10.50	81.64[k]	9.00	87.82[l]	14.23

Note: Reading tests results from: Wide Range Achievement Test; Woodcock Johnson – Revised; Wechsler Individual Achievement Test; Wechsler Individual Achievement Test – Second Edition; Woodcock Johnson – Third Edition.

FSIQ = Full Scale Intelligence Quotient

[a]*n* =28. [b]*n* = 28. [c]*n* = 32. [d]*n* = 31. [e]*n* = 40. [f]*n* = 37. [g]*n* = 23. [h]*n* = 24. [i]*n* = 36. [j]*n* = 22. [k]*n* = 39. [l]*n* = 39.

Table 4.22

Comparing Two-Way ANOVAs (IQ Classification X Gender) for Mathematics Calculation Utilizing Standard FSIQ and FSIQss

	Standard Full Scale IQ				Symbol Search Substitution FSIQ			
	(n = 196)				(n = 196)			
Source	*df*	*F*	*p*	*MSE*	*df*	*F*	*p*	*MSE*
Classification	2	17.25	.00**		2	10.16	.00**	
Gender	1	.39	.53		1	1.10	.30	
C X G	2	1.87	.16		2	.89	.41	
within group	190			139.09	177			146.15

Note. WIAT = Wechsler Individual Achievement Test. Classification = Borderline, Low Average, Average. C X G = Interaction of IQ Classification and Gender.

FSIQss = Full Scale Intelligence Quotient with Symbol Search Substitution.

** $p < .01$

Testing Hypothesis Four.

> *The fourth element of this study examined the effect of using an abbreviated IQ on subsequent classifications. Intellectual Abilities (IA) instead of the FSIQ was utilized to summarize WISC-III results. The Null Hypothesis assumes there will be no difference in mean levels of performance on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys when the IA, an abbreviated method of summarizing WISC-III results, is employed.*

On the surface, results from the second research question might suggest that by combining the VCI and POI all observed differences between the three IQ classifications would be preserved.

Utilizing Canadian norms provided by Weiss, Saklofske, Prifitera, Chen & Hildebrand (1999) the VCI and POI were converted to a Full Scale Intelligence Quotient (FSIQ) creating a slightly abbreviated eight subtest version of the traditional ten subtest

WISC-III FSIQ. Previously referred to as the General Ability Index, the name was changed simply to Intellectual Abilities (IA) in this study to emphasize the cognitive components of this standard score. Additionally, the IA deviates from the traditional IQ nomenclature in an attempt to more accurately depict what the WISC-III actually measures. Historically, the concept of IQ brandishes a storm of controversy in many academic circles. Although results from the present study will not likely alter perceptions of intelligence tests in general, there is a possibility that it could open the door to changing approaches to the way in which school psychologists interpret and report psychoeducational assessment information.

With the IA, the elimination of Arithmetic and Coding may have more of an effect if these subtests are areas of weakness or strength by subsequently raising or lowering the individual's FSIQ. A consistent theme in the first and third research questions has been the lower then expected levels of achievement in Mathematics abilities in the sample. This did appear to explain, at least in part, why the mean Arithmetic subtest performance for the Low Average group was not significantly different from that of the Borderline group. Therefore, a logical prediction might be that eliminating the Arithmetic subtest may result in more increases in IQ for the Low Average group than the Borderline and Average groups.

However this was not the case as Tables 4.23 and 4.24 clearly indicate that 12 increases in IQ classification occurred in the Borderline group with only nine increases in classification occurring in the Low Average group when the Intellectual Abilities method was employed. This represent an increase of four more Borderline individuals reclassified as Low Average as compared to the FSIQ Symbol Search substitution method of summarizing WISC-III results as previously illustrated in Table 4.18.

Table 4.23

Standard FSIQ Classifications Compared with Changes in Classification resulting from Intellectual Abilities

	Change in Classification after IA Replacement			
	No Change	Increase	Decrease	Total
Borderline	40	12	7	59
Low Average	51	9	8	68
Average	64	1	4	69
Total	155	22	19	196

Note. IA = Intellectual Abilities.

Table 4.24

Comparison of IQ Classifications between Standard FSIQ and Intellectual Abilities

	Classification According to Standard FSIQ			
IA Classification	Borderline	Low Average	Average	Total
Extremely Low	7	1	0	8
Borderline	40	7	0	47
Low Average	12	51	4	67
Average	0	9	64	73
High Average	0	0	1	1
Total	59	68	69	196

Note. FSIQ = Full Scale Intelligence Quotient. IA = Intellectual Abilities.

Following the same rationale utilized in the third research question, statistical analysis proceeded to determine if the abbreviated IQ represented by the eight subtest Intellectual Abilities (IA) version would produce the same results witnessed previously when academic achievement levels were compared to IQ classifications and gender. The 3 X 2 between subjects factorial design was repeated with results summarized in Tables 4.25 and 4.26. As expected, the shifting of a few individuals between IQ classifications did not radically change mean Reading and Mathematics achievement levels reported in the appendix (Tables A8 and A9). However, one obvious change was evident in the standard deviation of Low Average females' word reading abilities with considerably less variation in mean scores evident when the Intellectual Abilities method is used compared to the standard FSIQ. This pattern was not repeated when Low Average females' mean Mathematics scores are compared. Even with the observed change in word reading's standard deviation only occurring within the Low Average females group, Levene's test of Equality of Error variance across groups was not statistically significant.

Table 4.25

Comparing Two-Way ANOVAs (IQ Classification X Gender) for Word Reading Utilizing Standard FSIQ and IA

	Standard Full Scale IQ (n = 196)				Intellectual Abilities (n = 187)			
Source	*df*	*F*	*p*	*MSE*	*df*	*F*	*p*	*MSE*
Classification	2	17.25	.00**		2	7.68	.00**	
Gender	1	.39	.531		1	.29	.59	
C X G	2	1.87	.157		2	2.41	.09	
within group	190			139.09	181			156.85

Note. WIAT = Wechsler Individual Achievement Test. Classification = Borderline, Low Average, Average. C X G = Interaction of IQ Classification and Gender.

IA = Intellectual Abilities.

** $p < .01$

Table 4.26

Comparing Two-Way ANOVAs (IQ Classification X Gender) for Mathematics Calculation Utilizing Standard FSIQ and IA

	Standard Full Scale IQ (*n* = 196)				Intellectual Abilities (*n* = 187)			
Source	*df*	*F*	*p*	*MSE*	*df*	*F*	*p*	*MSE*
Classification	2	17.25	.00**		2	9.36	.00**	
Gender	1	.39	.53		1	.45	.51	
C X G	2	1.87	.16		2	.93	.40	
within group	190			139.09	181			153.22

Note. WIAT = Wechsler Individual Achievement Test. Classification = Borderline, Low Average, Average. C X G = Interaction of IQ Classification and Gender.

IA = Intellectual Abilities.

** $p < .01$

Similar to the first and third research questions, the null hypothesis for the fourth research question was also rejected for classification but not for gender when academic achievement levels (word recognition and mathematics calculation) were considered. More specifically, the classifications (Average, Low Average, Borderline) based on the IA method of summarizing WISC-III results served to differentiate individuals when reading and mathematics abilities were considered.

Following the trend set in post hoc analysis for previous research questions, the Average groups' mean academic achievement levels were significantly higher, than that of the Low Average and Borderline classifications. No statistically significant differences were found between mean achievement levels when the Low Average and Borderline classifications were compared. In contrast, males and females academic achievement levels did not differ significantly. An interaction effect between gender and IA-based classifications of Average, Low Average and Borderline was not evident.

Summary

Of the 196 archived files selected from the Education Clinic, 88 were females and 108 were males. Despite attempts to keep the numbers of females and males included in the study equal, several factors interfered with this goal. First and foremost, males referred to the clinic consistently outnumbered females. Secondly, while attempting to keep the numbers of Average, Low Average and Borderline individuals selected similar, both gender and age had to be considered as well. However, the main limiting factor in file selection was the elimination of all cases where statistically significant differences between Verbal and Performance Scales on the WISC-III were evident. This was considered to be an important selection factor given the study's goal of explicating both cognitive and academic profiles of individuals according to their IQ-based classification.

Although grade retention was only part of the descriptive information exhumed from selected files, it bears mentioning. Referred to as a 'retention-effect' in this study, 20% of the sample experienced a grade retention (males – 24% versus females – 15%). These retentions typically occurred more frequently among individuals with lower IQs as evidenced by the following breakdown: Borderline – 31%, Low Average – 21% and Average – 10%. The effect of utilizing age-based norms on standardized tests of achievement serves to accentuate known deficits of individuals with a history of academic failure. Calculating the 'retention effect" by simply subtracting the grade based standard score from the age-based is proposed as one way of reducing the inherent disadvantage placed on intellectually and academically challenged individuals. It is suggested that the validity of standardized measures of achievement are improved when individuals who have experienced grade retentions are assessed according to their present grade level as opposed to the standard set by their chronological age.

Correlation.

Significant positive correlations were found between the standard FSIQ, FSIQ (Symbol Search substitution), Intellectual Abilities (IA) and measures of word reading and mathematics calculation. This implies that as IQ increased proportionate increases in these achievement variables were also present in the full sample. Similar trends were also observed when female and male populations were considered separately. The removal of the Arithmetic subtest in the IA was associated with a lowering of correlation coefficient levels in Mathematics achievement. This might indicate that the Arithmetic subtest from the WISC-III is a confounding variable when utilizing the Full Scale Intelligence Quotient as a means of determining the presence of significant deficits in mathematics abilities.

Utilizing the standard method of IQ calculation the sample was subdivided into the three classifications (Average, Low Average, Borderline). With combined female and male groupings a significant negative correlation with Reading was obtained for Low Average individuals. Increases in IQ within the 80 to 89 range were significantly associated with corresponding decreases in levels of word reading ability. The opposite occurred with the Average group where a significant positive correlation occurred. Results for the Borderline group were not significant.

Research Hypotheses

The method of summarizing WISC-III results (FSIQ standard, FSIQss, IA) produced some changes in individual IQ classification. For example, Symbol Search substitution resulted in IQ classification changes as follows: either increase or decrease for 28% of Borderline, 20% of Low Average and only 7% of Average individuals in the sample. Symbol Search was a strength for 16% of the Low Average group where substantially more increases than decreases were evident. This same pattern was not repeated with either the Borderline or Average groups

Despite these slight alterations in the actual composition of the Average, Low Average and Borderline groups following the FSIQss and IA procedures certain statistical trends were consistent for the first, third and fourth research questions.

Hypothesis One.

(a) There will be no difference in performance between individuals on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys.

(b) There will be no difference in performance between individuals on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline individuals grouped as Younger (ages 6, 7, 8, 9 years), Middle (ages 10, 11, 12) and Older (ages 13, 14, 15, 16).

The first questions examined the effect of three independent variables (IQ classification, gender and age grouping) on standardardized measures of academic achievement (word reading and mathematics calculation). For both Reading and Mathematics achievement, IQ-based classifications served to differentiate the Average group from the Low Average and Borderline groups. However performance on these achievement measures did not result in the emergence of significant differences between Low Average and Borderline individuals in the sample. This trend was repeated in the third and fourth research questions. Therefore it might be assumed that the WISC-III FSIQ cannot be expected to reliably separate individuals into homogeneous groupings of academic aptitude unless a broader spectrum of standard scores are considered as evidenced in the Average IQ classification.

Unique to the first research question was the inclusion of the WIAT Only grouping of Average, Low Average and Borderline individuals. With the same statistical trends emerging with the WIAT Only subgroup (*n* = 103) as when all five measures of academic achievement were grouped together (*n* = 196), validity concerns pertaining to the obtained word recognition and mathematics calculation standard scores were addressed. Additionally Levene's test of equality of error variance was not significant for the WIAT Only or full sample.

An inverse relationship between age grouping and levels of achievement in mathematics was discovered. Unlike word reading ability that followed the expected age trend, calculation abilities decreased as the age of individuals increased. This was consistent across all three IQ classifications.

The interaction between independent variables on Research Questions One, Three and Four was not significant allowing interpretation of significant findings to proceed on a more solid footing.

Hypothesis Two.

> *Individual differences in factor indexes - Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), Freedom from Distractibility Index (FDI), Processing Speed Index (FDI) – were examined to determine if Average, Low Average and Borderline students could be differentiated. The*

null hypothesis states that there is, in effect, no difference between IQ classifications when corresponding WISC-III factor scores are compared.

The second research question examined the WISC-III factor scores in an attempt to determine similarities and differences between the Average, Low Average and Borderline groups. Mean levels of performance on three out of four factors, Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), Processing Speed Index (PSI), were statistically different when the three IQ classifications were compared. The Freedom from Distractibility Index (FDI) was the exception due to the fact that these differences were not found when the Borderline and Low Average groups were compared. This finding was detected by the Games-Howell post hoc procedure that is designed for when population variances differ. Similarities between the Borderline and Low Average groups were more obvious when analysis of the mean levels of performance on the two subtests (Arithmetic and Digit Span) that comprise the FDI was completed. Mean Digit Span scores were similar for Borderline ($M = 7.64$) and Low Average ($M = 7.85$) while the Average group's results ($M = 9.01$) fell as expected, into the average range. This was also the case for their Arithmetic results ($M = 9.01$) while the significantly lower Borderline ($M = 5.98$) and Low Average ($M = 6.78$) groups were not well differentiated by this subtest.

By accepting the null hypothesis for the FDI in the second research question it could be concluded that the Low Average group did not perform significantly better then the Borderline group. With Arithmetic being one of the subtests that forms the FDI, it is likely relevant to recall the findings from research question 1 (b) where it was found that achievement levels in Mathematics calculation declined as the age of the individual increased for all three IQ classifications. Therefore it could be concluded that a facility with basic mathematical operations is not apparent for many of the individuals in the sample. This being said, it is interesting to note that the null hypothesis was rejected when the Average IQ group was compared to both the Borderline and Low Average groups. Observed deficits in Mathematics achievement levels among the Average group did not translate into significantly lower performance on the FDI as was the case for the Low Average group in particular.

An additional step was added to the original research question to determine if gender might contribute to differential Factor score performance when IQ classifications were compared. Results were the same for the two main factors (VCI and POI) with mean differences found to be significant for all possible combination of classifications. The same pattern was found with the FDI results when gender was considered. The null hypothesis continued to be accepted for the Borderline and Low Average groups, even when the FDI for males and females were analyzed separately. However, gender did appear to have an effect on PSI performance. The null hypothesis continued to be rejected when only females' results were compared while it was accepted when just the males were considered. Significant differences were not found between males' PSIs belonging to the Borderline and Low Average groups. It should be recognized that the stability of these findings is related to the smaller sample sizes that result from subdividing the whole group according to gender and IQ classification. However, the number of males per classification always exceeded the number of females per classification. The lowest numbers were found among the female Borderline and Low Average groups ($n = 28$) with male classification groupings all exceeding 30.

Hypothesis Three.

This portion of the study introduced a second method of calculating the Full Scale IQ (FSIQ) by substituting Symbol Search for Coding. The Null Hypothesis assumes there will be no difference in mean levels of performance on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys when the Symbol Search substitution (FSIQss) method is employed.

Hypothesis Four.

The fourth element of this study examined the effect of using an abbreviated IQ on subsequent classifications. Intellectual Abilities (IA) instead of the FSIQ was utilized to summarize WISC-III results. The Null Hypothesis assumes there will be no difference in mean levels of performance on two measures of academic achievement (word recognition and mathematics calculation) among Average, Low Average and Borderline girls and boys *when the IA, an abbreviated method of summarizing WISC-III results, is employed.*

The third and fourth research questions were originally added to the study to explore the effect of changing the method of summarizing the WISC-III FSIQ on subsequent IQ based classifications. In doing so, this main goal was achieved but other questions emerged as a result of attempting to combine psychometrics with a more humanistic agenda. Namely, inevitable changes in classification did occur but mainly within the Borderline and Low Average groups. Figure 4.18 aptly illustrates the differential effect on IQ classification with the three methods of summarizing the FSIQ. If either the Symbol Search Substitution method or Intellectual Abilities abbreviated IQ method are utilized, a selected group of individuals will be reclassified. On a personal level, one can only imagine the emotional effect of being told your IQ classification has changed. Yet from a research perspective, the statistical significance of these changes was not found. By comparing the newly formed Borderline, Low Average and Average groups that resulted from reclassifications in Research Questions Three and Four with the findings from Research Question One it was hoped that some differences might emerge. Instead, the results were simply replicated. That is to say, reading and mathematics achievement levels for the newly formed Average group always were significantly better than the remaining classifications, while differences between the Borderline and Low Average groups were not found. Gender did not serve to differentiate IQ classifications according to obtained reading or mathematics achievement levels no matter what FSIQ method was employed.

classification and remaining classifications for both Reading and Mathematics achievement. However, when only gender was considered significant differences in achievement levels were not found. There was no significant interaction effect among these independent variables (IQ classification and gender).

Descriptive statistics for reading achievement are presented in Table 4.19 for each of the three IQ classifications according to the gender of the participant with both the standard FSIQ and Symbol Search Substitutions methods represented separately. Although some variation in mean word reading ability is observed when comparisons are made between these groups, there is little change in the variance of scores as reflected in the standard deviations.

Table 4.19

Comparison of Reading Achievement Standard Score Means for Standard IQ and Symbol Search Substitution Classifications According to Gender

	Standard Full Scale IQ					
	Borderline		Low Average		Average	
Reading (all tests)	*M*	*SD*	*M*	*SD*	*M*	*SD*
Female	86.79[a]	13.37	93.07[b]	12.19	99.66[c]	12.47
Male	90.13[d]	12.07	90.05[e]	11.80	101.08[f]	11.57
	Symbol Search Substitution FSIQ					
Reading (all tests)						
Female	88.52[g]	13.30	95.21[h]	12.19	97.39[i]	13.84
Male	90.64[j]	13.30	90.74[k]	11.79	98.38[l]	13.06

Note: Reading tests results from: Wide Range Achievement Test; Woodcock Johnson – Revised; Wechsler Individual Achievement Test; Wechsler Individual Achievement Test – Second Edition; Woodcock Johnson – Third Edition.

FSIQ = Full Scale Intelligence Quotient

[a]*n* =28. [b]*n* = 28. [c]*n* = 32. [d]*n* = 31. [e]*n* = 40. [f]*n* = 37. [g]*n* = 23. [h]*n* = 24. [i]*n* = 36. [j]*n* = 22. [k]*n* = 39. [l]*n* = 39.

Chapter 5

Discussion

Abstract

The present study attempts to provide some needed clarification of what, if any, differences distinguish individuals classified as Average, Low Average and Borderline according to their WISC-III FSIQ. This obtained or fallible score (Charter & Feldt, 2001) does not consider measurement error when establishing IQ classifications. As a result, a host of assumptions regarding the homogeneity of group membership could be made based solely on this classificatory label. Results from this study serve to highlight individual differences as well as similarities both within and between these three groups based on a random sample of 196 children between the ages of 6 and 16 who have completed psychoeducational assessments at the University of Alberta's Psychological Testing Center housed in the Faculty of Education.

Delimitations

The main intention of this study was to focus on a specific subpopulation of students as defined by their results on a commonly used psychometric instrument, the Wechsler Intelligence Scale for Children, Third Edition (WISC-III). There is no inherent belief that "Low Average" students possess any intrinsic qualities that would serve to differentiate them from students classified as "Average" or "Borderline" by the WISC-III. Instead, it may be more accurate to hypothesize that their similarities far outweigh their differences (Glutting & McDermott, 1994; Ward, Ward, Glutting & Hatt, 1999). Further comparative studies may choose to pursue an analysis of students from various IQ classifications on any number of academic and personality dimensions to determine whether more qualitative differences exit.

Having to make a decision in the absence of evidence is every clinician's worst fear. This same concern guides the current research study. Many psychologists use an assessment interview approach when administering standardized tests to individuals. During the course of the assessment, they may have several opportunities to obtain valuable information about the client's feelings and thought processes, including evidence of logical and creative thinking as well as their ability to make both verbal and nonverbal associations. Even in the context of establishing rapport with the client, trust and assurance is built that the information gained from the assessment will be combined with other sources of supporting data to create an exploratory profile of the individual. This element of trust implies that when all sources of information are combined and reported in an accurate manner, there will be no resulting harm for the client. Every effort is made to validate and triangulate information within the practical constraints of time and available resources. The use of archival data assumes that these same professional standards have been adhered to during the assessment process. Additionally, some information about the client might be lost when only the test protocols and psychoeducational report is being accessed. Some of the more qualitative information gained during the actual assessment could support the development of an overall profile for groups of individuals from the same IQ-based classification. This type of study would seem like a logical sequel to the present one.

Academic achievement data for the study were limited to only two subtests gathered from six different standardized measures. It could be argued that including a reading comprehension task might have revealed more information than simple word recognition. However, there is considerably less consistency in the manner in which an individual's comprehending abilities are measured among various tests of reading as compared to the more straight forward identification of isolated words. Similarly, the mathematics calculation task included in the study would produce more consistent data than various forms of reasoning or problems solving. Therefore, it is recognized that reading and mathematics achievement levels reported in the study may be considered basic but hopefully more consistent, if not reliable, when the data obtained from the various standardized instruments is combined and compared.

Recognition of comorbid conditions such as Learning Disabilities or Disorders of Behavior and Attention would be secondary to the actual research goals. This information would be used primarily for descriptive purposes. It could be argued that including such students would introduce a series of confounding variables. On the other hand, it may be difficult to obtain research participants who are completely free of any learning or behavior disorders, given the nature of the targeted student population. Despite the availability of professional and academic standards for the identification of disorders of attention, learning, and behavior in children, there remain some inconsistencies in the actual diagnostic processes for these conditions. During the ten year span that the data for this study covers, there have been a number of research related changes to the diagnostic criteria, not to mention changes in theoretical frameworks that have impacted school psychologist approaches to these rather complex conditions. Therefore, when individuals with learning disabilities and attentional or behavioral disorders are reported in this study, the rather tenuous nature of these labels is recognized. No attempt was made to form a causal link between these conditions and specific IQ-based classifications other than simply reporting trends that appeared to emerge from reviewing the 196 assessment files included in the study.

Rationale Revisited

School Psychology is the embodiment of an interdisciplinary area of study. When two distinct yet inter-related specializations, education and psychology, are combined the result may appear to be somewhat confusing and certainly not without its share of controversies. At the same time, individuals who are drawn to this field may recognize its inherent practical applications and front-line possibilities.

Utilizing theory to fuel the engine, the power of knowledge provides the necessary torque that allows school psychologists to navigate through a variety of terrains and climates. Ever sensitive to the needs of students, families, and educators, the professional paths that school psychologists might take are often filled with the proverbial twists and turns. It is incumbent upon professional practitioners to remain focused on their destination, while remaining mindful of the detours and potholes along the way. Research helps to map this course and creates signs and symbols to mark significant findings as well as post warnings of impending hazards. If ignored, the school psychologist could drift aimlessly amid a myriad of referrals and assessments, never questioning the status quo or entertaining thoughts of paving new directions for their profession.

The present study attempts to survey the psychometric landscape. As a result of the analysis of a decade of psychoeducational assessments conducted at the University of Alberta's Education Clinic, an archival map composed of 196 individual profiles has been formed. Can this historical information influence current practices and decision making in school psychology? This is not a completely rhetorical question. Its answer lies in the hearts and minds of those school psychologists who continue to utilize today's revised and renormed tests of intelligence on a regular basis. Albeit new and improved, these psychological instruments continue to reinforce basic assumptions surrounding the measurement of individual differences. Any IQ-based classifications that might result will be reported in subsequent reports and continue to reinforce if not legitimize educational decisions surrounding programming and expected outcomes for the students being assessed. In other words, this would-be psychometric makeover could be just another wolf in sheep's clothing for Borderline and Low Average individuals who are attempting to succeed in educational environments more attuned to the mainstream or Average students. Subsequent research might reject this hypothesis in favor of a more positive prognosis for the efficacy of psychoeducational assessment in today's schools. To this end, history can help guide the future, just as the results from this study could influence attitudinal changes, if not policy revisions.

The present study contributes to the ongoing quest of school psychologists to better understand and utilize results from intelligence tests by attempting to identify specific markers that might serve to differentiate students according to their IQ-based classifications.

Calculated Differences.

If logic and empirical truths are used as guides, one might hope that the resulting statistical proofs may indeed serve a valuable purpose in the overall quest for knowledge. At times considered a second order purpose (Field, 2000), educational research can advance knowledge by refuting or at least calling into serious doubt previous assumptions and practices that continue to impact students, teachers, and educational systems. As a case in point, current assessment practices employed by educational psychologists frequently utilize standardized measures of intelligence such as the WISC-III as a means of classification, categorizing and coding school-age children and adolescents. Frequently, the main reason for involving a school psychologist in the assessment process is to provide information concerning the student's cognitive functioning. Establishing estimated IQ levels will in turn determine if government funding for students with special needs can be accessed. Sattler's (2001) four pillars of assessment are often reduced to one or possibly two elements, with standardized test results or IQ often carrying the most weight or political punch. When cast into the same category as phrenology and physiognomy, the measurement of IQ appears to be the loan survivor in the game of human classification (Gifford & O'Connor, 1992).

If the main goal of assessment is the remediation of problems with learning (Sattler, 2001) then the classification of students according to estimated levels of intelligence may indeed be a second order purpose. The unique educational needs and characteristics of the individual cannot be readily determined solely by referring to standardized test scores, no matter how psychometrically sound the instruments may be. Sattler advocates caution when employing classification systems: "In developing assessment findings and recommendations, be guided by the child's performance and not by a classification system of arbitrary cutoff scores on an intelligence test" (p. 26). Yet at the same time Sattler goes on to recommend adhering to a classification system's specific cutoff points and labels when reporting test results. This means that a test instrument's accounting for measure error through the utilization of confidence intervals is not considered when reporting the precise classification of an aggregate test score such as the Full Scale Intelligence Quotient (FSIQ) on the WISC-III.

According to psychometric theory, an individual's 'true score' includes measurement error (Nunnally & Bernstein, 1994). As a hypothetical construct, the actual 'true score' cannot be observed (Sattler, 2001). As Charter and Feldt (2001) point out, "knowledge of the examinee's true score is available only when the test score has perfect reliability, which none of our tests do" (p. 350). Charter and Feldt refer to an individual's true score as a fallible score and warn that it should be treated as such. They remind examiners of the inherent inexactness of the obtained score and advocate the use of confidence intervals to indicate a range of "possible true scores" (p. 350).

It is assumed that some traits, such as intelligence, are stable enough that they can be measured reliably. Yet when classifications systems based on cut off scores are introduced into this rather tenuous measurement process, it could be argued that the presence of error is being ignored if not forgotten. This seemingly flagrant violation of basic psychometric theory is so commonly practiced that school psychologists involved in psychoeducational assessment rarely question it. Consumers of psychoeducational reports expect, if not demand, classifications or diagnostic labels to ensure that applications for funding will be accepted by government auditors. But the "end justifies the means" argument has worn too thin to withstand public scrutiny. An individual's right to needed educational services without his or her being categorized with diagnostic labels is on the horizon of a growing human rights movement (National Association of School Psychologists, 2004).

The present study attempted to provide some needed clarification of what, if any, differences distinguish individuals classified as Average, Low Average and Borderline according to their WISC-III FSIQ. Together, cognitive and academic achievement profiles served to outline individual strengths and weaknesses. Statistical analysis helped determine whether these observed differences and similarities can be extrapolated from the sample's 196 Average, Low Average and Borderline children to the general population. These classifications might be refuted as statistically unsound due to such factors as measurement error and confidence intervals that result in a blurring of clearly demarked boundaries. Measurement theory aside, IQ-based cutoff scores and classification systems continue to have a significant impact on the school psychologists' role as well as the individuals being served through psychoeducational assessment practices.

As a fledgling researcher, one is humbled by the daunting task of attempting to make a meaningful contribution in such an important yet often misunderstood area of study. Among the plethora of theories of intelligence and related research, the present study takes a somewhat atheoretical stance and centers its philosophical energy on such practical concerns as the true meaning of IQ-based classifications. Far from being mundane, the real-life implications of being labeled and grouped according to estimated levels of intelligence have both immediate and long-term consequences. Realistically, the present study will not change this practice but, at the very least, plant seeds of doubt surrounding the efficacy of utilizing these classifications as an administrative tool to justify the allocation of special needs funding and academic supports as well as special class placements.

Retention Effect – A Psychometric 'Fly in the Ointment.'

Close to 20% of the sample experienced a grade retention, with individuals possessing low IQs (Low Average and Borderline) reporting more retentions than those with Average IQs. On the surface, this occurrence could seem unexpected, in that decisions to have individuals repeat their grades occur more frequently when low ability students struggle to meet curricular and academic standards. However, often overlooked is the fact that as these children continue to progress through school, the lingering consequences of being retained extend beyond any possible psychological effects that the student may have experienced. 'Retention effect' was used to describe the residual psychometric influence of being continually compared to one's age peers on standardized tests when these students are not at the same grade level. The 'retention effect' not only exacerbates known differences in aptitude and academic achievement levels but also serves to create artificial discrepancies between them. The following provides a brief overview of some of the common features and possible implications of the 'retention effect':

- The psychometric equivalent of the Matthew Effect (Stanovich, 1986), which is activated through the application of age norms on standardized test. Known deficits in academic achievement will be exaggerated when a student's age and corresponding grade no longer match expected trends.
- Calculating the 'retention effect':
  When students have experienced grade retentions or participate in decelerated programs, the simple difference between grade and age norms in standardized measures of achievement can be calculated. Actual progress in learning can be more readily measured if a more realistic approach is taken.
- The curriculum advantage of non-retained students is ignored. In addition, individuals who are in modified or alternative programs may not follow standard curriculums, which may place them at a further disadvantage when normative tests of achievement are administered; same age peers who follow regular curriculums may have a distinct advantage when given normative tests of achievement irrespective of individual differences in cognitive abilities. In other words, they have a curriculum advantage.

The 'retention effect' is just one more consideration in the assessment of students experiencing academic challenges. Rather than expecting the struggling student to achieve at an artificially higher level, it is more realistic to consider their grade peers as the legitimate comparison group to gauge academic progress.

Correlations of Interest

Moderate but significant correlations in the present study reconfirmed the known connection between IQ and academic achievement (Kaufman, 1994; Sattler, 2001). The method of determining the IQ, standard FSIQ, FSIQ with Symbol Search substitution or the abbreviated IA, all produced similar correlational results with the full sample and within gender groupings. Although some increases and decreases in IQ did result from the two alternative methods of calculating IQ, these changes were not sufficient to affect overall associative trends in word reading and mathematics calculation abilities when specific IQ classifications were compared.

However, one interesting and unexplained result did occur when the Average (n = 69) and Low Average (n = 68) groups' correlations on a measure of their reading skills were compared. Their low but significant correlations were similar in strength but not direction (r = .24 for the Average group; r = -.26 for the Low Average group). In both cases, roughly 6% of the variance in individual word reading ability is explained by their respective IQ classifications. Therefore, it is safe to say that IQ-based classifications offer only limited explanatory power with respect to possible factors responsible for success on these word identification tasks. Yet the negative relationship within the Low Average group was not observed within the Borderline group (n = 59) where no significant correlation was found.

The Low Average groups' significant negative correlation with this achievement variable does warrant further investigation. The word reading task in question is devoid of contextual clues. Is it possible that Low Average readers may rely more heavily than their Average peers on contextual clues to recall individual words? More importantly, what is unique about the processes involved in a word recognition task that serves to challenge the Low Average student? This significant negative association within the Low Average classification implies that as IQs increased, word reading ability decreased. This same pattern was not repeated within the Low Average classification when levels of mathematics achievement and IQ were compared. Only the Average classification revealed a low but significant positive correlation with mathematics calculation tasks.

The word reading task included in this study was simple and basic, but it has been argued that the ability to read words quickly, accurately, and effortlessly is a critical skill in reading (Adams, 1990; Downing & Leong, 1982; Stanovich, 1994). At this early 'mastering' or 'automaticity' phase of learning to read, overlearning of the individual letters, spelling patterns, pronunciations, and meanings of words is required. As Stanovich points out, the demands of inefficient word processing skills leave fewer cognitive resources available for the higher level processes of text integration and comprehension. Both cognitive and linguistic processes are enacted to comprehend the meaning of the text's message. The process of comprehension depends on the reader actively searching for the overlap in meaning between words and phrases. The detection of syntactic and semantic coherence depends critically on the automaticity of word recognition (Adams). Words "that are highly familiar will be mapped directly, instantly, and effortlessly from sight to meaning" (Adams, p.413). Therefore, the speed and accuracy with which readers can respond coherently to the orthographic, phonological, and semantic patterns of text will determine reading proficiency and support comprehending processing.

Evans, Floyd, McGrew and Leforgee (2001) demonstrated that a moderate association exists between processing speed, basic reading skills, and reading comprehension for individuals between 6 and 10 years. They hypothesized that "the more rapidly and efficiently an individual can automatize basic academic or cognitive operations, the more attention and working memory resources can be allocated to higher level aspects of task performance" (Evans, Floyd, McGrew & Leforgee, p. 257). Therefore, a logical follow-up to the present study would be to investigate, in greater detail, the reading comprehension abilities of Low Average individuals as compared to their Average and Borderline peers.

Speed of Processing and Intelligence

The Processing Speed Index (PSI) is comprised of two WISC-III subtests, Coding and Symbol Search. Within all three IQ classifications, the mean Symbol Search score exceeded the mean for Coding but remained less than a two point scaled score difference. Prifitera, Weiss & Saklofske (1998) suggested that when a difference of five points or more exists between Symbol Search and Coding, the PSI would not be considered valid.

On average, this was not the case for the present sample although it is recognized that for 20% of the cases, differences between these two subtests was significant. Slightly more variability in Coding and Symbol Search results was observed in the Low Average and Borderline groups but this difference was not considered extreme due to the fact that it was within a one point difference when compared to the Average group's standard deviation for these separate subtests. When significant differences between Coding and Symbol Search exist, Prifitera et al. point out that "the PSI composite will have little intrinsic meaning, and should not be interpreted as a unitary construct" (p.31).

For both the Borderline and Low Average groups, the mean PSI exceeded expectations and fell in the Low Average and Average ranges respectively. This trend did not occur with the Average group for whom the mean PSI fell, as expected, in the Average range. It has been suggested that strength in processing speed could facilitate learning through the efficient comprehension of novel information (Prifitera et al., 1998). Sattler (2001) described the hypothesized processing speed abilities that underlie this factor as a product of both perception and quickness of mental processing, "Processing speed measures the ability to process visually perceived nonverbal information quickly, with concentration and rapid eye-hand coordination being important components" (p. 233). Typically, success with tasks involving speed or quickness of mental processing is not associated with lower levels of cognitive functioning. An anomaly appears to exist where the current results fail to support fully this assumption, as evidenced by the better than expected mean PSIs for the Borderline and Low Average groups. Even though it is recognized that highly restrictive time limits can positively skew test results (Nunnally & Bernstein, 1994) this was not the case for the PSI, for which a two minute time limit is imposed for both Coding and Symbol Search. In addition to processing speed, other abilities measured by the PSI include attention, concentration, short-term visual memory, visual-motor coordination, cognitive flexibility, and fluid ability (Sattler, 2001).

Substituting Symbol Search for Coding had an effect of increasing the FSIQ and subsequent classification of 16% of the individuals included in the sample. When the productive element of the Coding subtest was removed, these individuals were able to employ their abstract thinking skills more effectively and efficiently. This increase in

FSIQ and classification was most evident in 16% of the Low Average group when compared to the Average and Borderline groups.

Cooper (1995) suggested that differential performance on Symbol Search and Coding is a reflection of underlying differences in the nature of these related tasks. Symbol Search requires mainly visual processing, while Coding "permits verbal encoding, possible associative learning, and a more extensive fine motor output" (p. 229). Clerical speed and accuracy, combined with pencil and paper skills, also serve to differentiate these related yet dissimilar tasks. It has been hypothesized that a host of factors may readily influence performance on Coding, and two of these stand out as being unique to this subtest. Concern for accuracy and detail, while attempting to quickly reproduce the targeted symbols, is eliminated in Symbol Search in which a simple yes or no response is all that is required.

Sustained effort, persistence, and motivation are required for success with both Coding and Symbol Search (Sattler, 2001). Yet subtle but significant differences in these related measures of processing speed could provide a window into how instructive materials, could be designed to enhance the performance of Low Average and Borderline individuals in the classroom. As a result of a reduction in the response complexities of the tasks, these individuals might be more able to apply their abstract reasoning skills. In addition, presentation of novel, as opposed to redundant stimuli, served to enhance nonverbal reasoning abilities, as witnessed by improvements in FSIQ levels following the Symbol Search substitution. Too much repetition can invoke the use of strategic memory, which may serve to overload already taxed cognitive processes. Fewer resources may remain in reserve to address the original task demands when processing abilities are split between several augmenting options. In other words, keeping the response simple without compromising on the cognitive complexities of the task could be an important element in the design and implementation of instructive materials for Low Average and Borderline individuals. This suggestion is consistent with current conceptualizations of working memory as "a limited capacity system allowing the temporary storage and manipulation of information necessary for such complex tasks as comprehension, learning and reasoning" (Baddeley, 2000, p. 2).

An episodic buffer is the fourth component of the Working Memory model proposed by Baddeley. It is described as the "process or mechanism for synergistically combining information from various subsystems" (p. 9). With an emphasis on the integration of information, Baddeley suggests that the episodic buffer acts as an interface between memory and conscious awareness. Previously, Baddeley and Hitch proposed a three component Working Memory model (Baddeley & Hitch, 1977), consisting of "an attentional control system, the 'central executive', aided by two subsidiary slave systems, the 'phonological loop' and the 'visuospatial sketchpad'" (p. 3).

With the current emphasis on integrating as opposed to isolating information (Baddeley, 2000), the complexities of the executive control function in Working Memory might be more clearly understood. Although the WISC-III's Freedom from Distractibility Index (FDI) is more commonly associated with Working Memory (WM) (Sattler, 2001), it should be noted that elements of the Processing Speed Index (PSI) are equally dependent on WM. In comparing the Coding and Symbol Search subtests that comprise the PSI, it is apparent that the demands on the visuospatial sketchpad might be greater on the former task, with kinesthetic and reproductive elements significantly reduced on the later.

The fact that substituting Symbol Search for Coding had a differential effect for each of the three IQ classifications may also have some instructional implications. For 16% of the Low Average group, their classification changed to Average following Symbol Search substitution, with only 4% of this group's classification falling to Borderline. While 14% of the Borderline group's classification changed to Low Average, 15% of this group's classification changed to Extremely Low. The effect of substituting Symbol Search was close to negligible for the Average group, with only 6% increasing their classification to High Average, while only 1% fell to Low Average. Clearly, retaining Coding in the FSIQ calculation benefited more Borderline individuals than it did for the Average and Low Average groups combined.

From a psychometric viewpoint, Symbol Search may be viewed as a purer measure of psychomotor speed than Coding according to Prifitera et al. (1998) due to the fact that it is "not affected by associative learning" (p. 32). Coding, on the other hand, involves the pairing of symbol content with numbers, and therefore deficits in short-term visual memory of these paired associations might influence performance. The rapid, error-free processing of what appears to be simple information has been connected to other types of routine tasks such as the information processing involved in reading (Prifitera et al.). Indeed, there may be some connection between the number-symbol deficits observed in Coding and the sound-symbol deficits in word recognition found more commonly among the Low Average group in this study. Further research with Low Average individuals is needed before blanket statements can be made pertaining to processing speed preferences although the present study does hint that one may indeed exist.

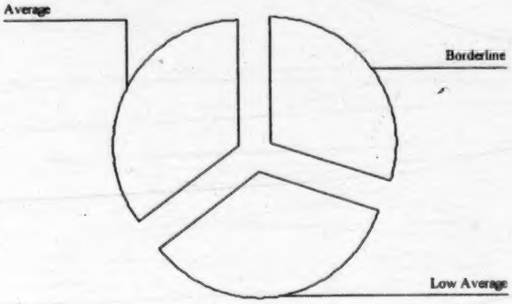
According to Dunn (2001), "sample size and the strength of a given correlation are actually independent of one another" (p. 225). Larger sample sizes do not necessarily result in larger correlations. Dunn adds that spurious correlations can be reduced if correlational analysis is performed with samples sizes comprised of a minimum of thirty participants. In the present sample ($N = 196$), a moderate association ($r = .45, p = .00$) between Symbol Search and Coding did exist. This relationship increased substantially ($r = .74, p = 00$) when cases in which five points or greater differences between these subtests were eliminated ($n = 154$). Another interesting trend emerged when FSIQ classifications were considered. A clear relationship was evident with the Borderline ($r = .36, p = .01$) and Low Average ($r = .46, p = .00$) groups with sample sizes of 59 and 68 respectively. This significant relationship was not the case for the Average group ($r = .17, p = .16$) with a sample size of 69. When the scatterplots are examined for each classification, it is evident that a linear relationship does not clearly exist between Symbol Search and Coding for the Average group, while it is more evident for the Borderline and Low Average classifications.

Finally, one last consideration pertains to connecting the present research findings with future measurement of the PSI on the WISC-IV. Despite several structural and cosmetic changes to this instrument, the Processing Speed factor remains relatively unchanged from its WISC-III counterpart. The core tasks, Coding and Symbol Search, remain the same, with only the administrative instructions being shortened and the number of items increased from 45 to 60 on form B of Symbol Search. Cancellation has been added as a supplementary subtest that can be used as a substitution for one of the

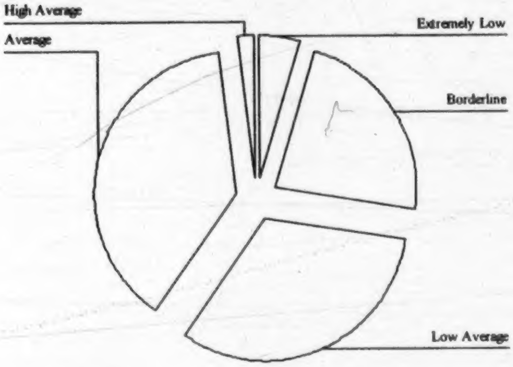
Figure 4.18 IQ Classifications According to Method of Summarizing WISC-III Results

Student Classification According to Full Scale IQ

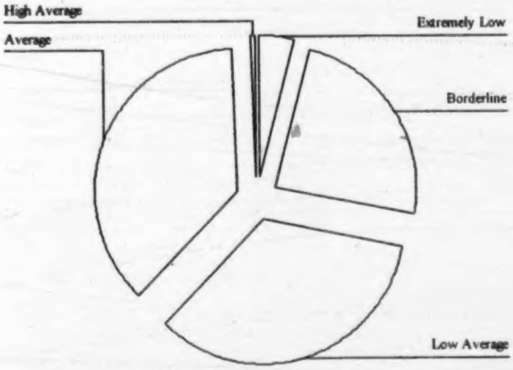
Standard Calculation



Student Classification According to Modified FSIQ



Student Classification According to Intellectual Abilities



knowledge, the implications is that they are able to remain focused on a task while ignoring imminent distractions. In his reference to expert versus novice performance on tasks judged to be cognitively demanding, Nickerson does not recognize individuals with significant attentional difficulties. Loarer (2003) indicates that "the application of metacognitive strategies consumes large amounts of attentional resources, the capacity of which is also limited" (p. 256). Is it merely assumed that mediated experiences aimed at building these metacognitive skills and abilities will adapt fully to the individuals' needs or will disorders of attention circumvent the expert-building processes such that it becomes an illusive, theoretical goal rather than an achievable one? Recalling the previous chapters' descriptive data pertaining to reported attentional difficulties within the research sample, it was clear that as IQ levels decreased, incidents of significant problems with sustained attention increased. Although, no standard criteria were established for identifying the nature and severity of these difficulties, suffice it to say that these 'problems with attention' were significant enough to warrant acknowledgement in the graduate-student-generated psychoeducational report. In many cases, these attentional difficulties were one of the reasons that the referral for an assessment was initiated. Measurement of intelligence and achievement is not only exacerbated but at times, invalidated when attentional deficits are severe and persistent. According to Carroll (1993):

> It can be argued that attention is involved, in varying degrees, in all cognitive performances and thus in all performances that are regarded as indicating cognitive abilities. One can expect it to be very difficult to separate the attentional components of such performances from those components that represent latent traits of abilities other than the ability to attend. An individual difference factor could often be equally well interpreted either as a factor of some particular cognitive ability or as a factor of attentional ability (p. 547).

In other words, with roughly 20% of the cases included in the research sample, IQ and achievement results may have been influenced negatively by reported problems with attention. Although confidence intervals may account for this form of measurement error, IQ-based classifications and government generated IQ-cut offs for special needs funding continue to ignore this common source of individual variability in test performance.

One theme in applied research on disorders of attention centers on the need to augment observed deficits in executive functioning through careful and explicit structuring of the individual's academic (Loarer, 2003) and assessment (Schwean & Saklofske, 1998) environments. The expectation is that all individuals can benefit from these targeted interventions, with IQ rarely, if ever, mentioned. Given the observed differences in WISC-III profiles for each IQ classification according to the Freedom from Distractibility Index (FDI) and such attention-laden tasks as Arithmetic and Digit Span, it is little wonder that Sternberg (2004) suggests that reasoning and intelligence are so interrelated that one can not be studied without reference to the other. Nickerson (2004) takes this one step further when he states, "There can be little doubt that intelligence, as measured by conventional IQ tests, is a reasonably good predictor of success both in academic performance and in the world of work." (p. 416). Although the jury may still be out on judging whether the curiosity and inquisitiveness of Low Average and Borderline individuals are sufficient to overcome apparent cognitive processing shortfalls, further research could utilize findings from the present study to, at the very least, develop differentiated training and modeling paradigms that more closely align with known attentional deficits commonly found within these IQ groupings. The cognitive and metacognitive approaches to instruction suggested by Knight, Paterson and Mulcahy (1998) in their Strategies Program for Effective Learning and Thinking (SPELT) program are one example of how research and practice can unite to produce supports for teachers, parents, and students:

Practicing school psychologist need to go beyond Nickerson's (2004) assumption that high IQ is predictive of academic and work related success when they consider the needs of Low Average and Borderline individuals. A reference to specific IQ levels is notably absent in the Vygotskian theory of child development. The dynamic nature of learning is encapsulated in Vygotsky's (1978) statement:

> Learning is not development; however if properly organized learning results in mental development and sets in motion a variety of developmental processes that would be impossible apart from learning. Thus, learning is a necessary and

> universal aspect of the process of developing culturally organized, specifically human, psychological functions. (p. 90)

Learning, according to Vygotsky, in a school setting can awaken intellectual processes and therefore it is incumbent upon the classroom teacher to create an environment where this is indeed possible. Again, there is no reference to limit setting IQ-based classifications or cut off scores beyond which hopes for academic improvements, and "learning" are diminished.

Ceci (1996) proposed that one advantage of utilizing the ZPD is that "children are exposed to the complete task while only engaging in those aspects found at the limits of their own cognitive competence" (p. 96). According to Vygotsky (1978), if children are unable to imitate a task then it is not within their developmental level. This can serve as a basic guide for teachers attempting to incorporate the ZPD within the confines of their classrooms and curriculums. More specifically, if observed deficits in mathematics abilities in the general student population are similar to the Low Average students' academic deficits reported in the present study, then there is an opportunity to utilize the ZPD to enhance their learning and developmental processes.

In keeping with the intentions of today's democratic classroom, Vygotskian principles can be applied effectively and successfully if the following is kept in mind: "human learning presupposes a specific social nature and a process by which children grow into the intellectual life of those around them" (Vygotsky, 1978, p.88). Planting seeds for learning requires more than a fertile knowledge base and skillfully scaffolded Individual Program Plans (IPP). Without engaging and supportive teachers and accepting classmates who help create an "intellectual life" that is equally nurturing and reaffirming, then the main environmental variables for learning will be absent. Loarer (2003) emphasized the mediatory role of the teacher when he states that it is "essential in cognitive education methods" (p. 248). Additionally, the energizing force of the ZPD can help create and sustain the "intellectual life" required for the Low Average and Borderline students to take full advantage of the assistance that is being offered.

Theory has a tendency to overlook some practical considerations at times. Within the present context of working with lower IQ students to ensure successful learning experiences in school, it is important to remember the most important variable. The students in question must be involved in the planning process to help ensure their commitment and accountability. This prerequisite may seem obvious, but it is one that is often ignored in many academic settings. How can one impose a ZPD on a reluctant, unmotivated, and lethargic participant? In these cases, students are often perceived as the problem rather than the planning process. Depending on their age and developmental level, students' and parental involvement can make the difference between success or failure of IPP goals and curricular adaptations or modifications.

Today's school psychologist can play a vital role in addressing the psychological as well as the academic needs of the student being assessed. Naturally, doing so requires viewing the student in question as much more than an accumulation of tests scores or estimated IQ level. An effective and carefully crafted psychoeducational report can help create a climate necessary for effective and progressive planning to take place. Keeping in mind that traditional static measures of intelligence test in the zone of actual development (Guthke & Beckman, 2003) the door is left open to expanding sources of information pertaining to the individual's cognitive abilities when testing based on the ZPD is permitted. This form of dynamic assessment can be seen as an alternative method of "uncovering the differences between latent capacities and developed abilities" (Guthke & Beckman, p. 239). In keeping with the spirit of this alternative form of measurement, school psychologists might utilize their 'testing the limits' option more frequently to obtain similar information on student cognitive potentials and in vivo processing abilities. In doing so, standardized procedures are not being violated and normative comparisons can still proceed.

This section will be conclude with some final theoretical considerations relating to environmental and individual variables germane to working with students from all IQ classifications within the classroom setting. Specifically, how can the ZPD (Vygotsky, 1978) and proximal processes (Ceci, 1996) be differentiated? Are these concepts one in the same or do substantial differences exist that may have educational implications?

Perceived as the "engines that drive development" proximal processes, are described by Ceci (1996), as the actualization of "genetic potentials for intelligence" through "specific mechanisms of organism-environment interaction" (p. 244). Similar to the ZPD, effective proximal processes will become progressively more complex as the specific needs of the individual change as a result of successful learning experiences. Perhaps genetically driven influences are presumed within the social interactions that are a hallmark of the ZPD. With both Vygotsky and Ceci the quality and nature of the interactions that occur are key to unlocking some of the mysteries that surround the dynamics of specific learning processes as well as general intellectual development. The main message for educators and school psychologists alike is to remain cognizant of the vital role that numerous environmental variables play within the "myriad of genetic potentials for intellectual behaviors" (Ceci, 1996, p. 245). Whether it is ensuring high levels of proximal process in the classroom or simply fine tuning the ZPD to match the present learning needs of the individual, a clear path for success has been paved by these visionary social scientists.

The extent to which school psychologists employ the four pillars of assessment (Sattler, 2001), might determine the degree to which standardized test results are contextualized within the broader spectrum of dynamic classroom interactions. If the same time was spent observing the student within their classroom milieu as was administering the test battery, then more credence might be given to this nonstandardized data. Individual test performance in a quiet, isolated room might not shed light on the student's test-taking strategies within the classroom. Do the reported problem with learning lie mainly within the student or their learning environment? Obviously both contribute to the student's actual performance in school and effective psychoeducational assessment reports reflect this reality.

In a broad sense, "cultural context is an integral part of cognition because the culture arranges the occurrences or nonoccurrences of events that are know to affect cognitive developments" (Ceci, 1996, p. 95). Similarly, the ZPD defines mental development in the process of maturing such that the nature and quality of social interactions and actual classroom climate in which they occur are vital components that must be acknowledged and nurtured. Today's highly trained curriculum specialist may need to be reminded that students' cognitive development is directly related to "the quality of social mediation" (Loarer, 2003, p. 248) present in the classroom, not just the quality of instruction.

Theory, Research and Practice Connections

According to Dawson and Guare (2004), "The goal of assessment is intervention." (p. 25). School psychologists, in particular, are well aware of the need to unite theory and practice in a seamless fashion such that their assessments, reports and recommended interventions fit not only the needs of the student, but the complexity of the educational milieu in which they are placed. So too, research in school psychology balances theoretical dictums with perhaps more mundane, yet essential, educational policies and guidelines.

While the current study created an archival map of some of the cognitive and achievement variables associated with Average, Low Average and Borderline students, the door remains open to more intense and integrated studies that link developmental theories of intelligence with differential performance on psychometric measures of intelligence. Perhaps the present study could be viewed as more exploratory in nature, given the fact that classificatory labels are rarely referenced in the research literature, yet are a common mode of communication in the school psychology field.

The ongoing debate over definitions of intelligence and its measurement seems endless and, on the surface, might appear to have little impact on the practitioner's daily task of providing professional assessment services. When school psychologists turn to current psychological studies for guidance and assistance, they might be frustrated by the lack of consensus among theorists. It is obvious that theoretical allegiances fuel the assessment industry, its related research and test development. Decades of disagreeing among social scientists is as much a tradition as the utilization of David Wechsler's tests in psychoeducational assessment batteries among school psychologists. Frustration can lead to resignation if clear directives do not emerge from their academic banter and sideswiping.

On of the goals of the present study was to fill the gap between pure academic research and the every day realities of school psychology practice. The malleability of an individual's IQ and their corresponding classificatory labels was examined for several theoretical and practical reasons. A clear demonstration of these changes in classification did not appear to affect overall trends when academic achievement levels were compared across the three IQ-based classifications. Therefore the question remains, are these changes in global scores meaningful or simply an addendum to known sources of measurement error? If cognitive processing strengths matched with selected psychometric tasks, such as Symbol Search for example, then positive outcomes resulted for those individuals. The Low Average group seemed to benefit most from this simple, yet significant procedural change. The very concept of individual differences implies that a range of strengths and weaknesses in intellectual abilities should exist. Further research with the Low Average population might examine both verbal and nonverbal processing capabilities with the intent of developing a more detailed picture of common strengths and challenges.

The concept of IQ is often associated with rigorous debate. Historical conceptualizations can be traced back over a century and certainly add spice, if not clarity, to ongoing applied research. It appears that Spearman's (1904) much maligned hierarchical model may have been misinterpreted as witnessed by Ceci's (1996) statement that "the traditional conceptualizations of intelligence are inadequate, especially the notion of a single underlying processor, *g*, that accounts for a substantial amount of interest variance and is innately determined and immutably fixed" (p. 72). Ceci goes on to provide a section of Spearman's infamous research article on 'General Intelligence' but chooses to omit a lengthy portion of the specific summative conclusion in question. By doing so, Ceci was able to support his own theoretical framework housed in a bioecological model. In reality, Spearman did acknowledge that cognitive tasks may have more than just a unitary factor or *g* in common when he stated "all branches of intellectual activity have in common one fundamental function **(or group of functions)**, whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others" (p. 284). The door was left open to the possibility of several functions combining to account for similarities between these various activities judged to be indicators of intelligence. Sternberg (2000) dismissed all arguments pertaining to *g*, "because the general factor applies to the narrow concept of intelligence but not to the broad one. Thus whether there is a general factor depends on how broadly one chooses to define intelligence." (p. 635).

If researchers in the field of intellectual measurement step outside of their theoretical boxes long enough, more similarities than differences in their views may become apparent. All attempts at imposing a structure on the epigenesis of intelligence is circumvented by these theorists' own cognitive limitations. Researchers and theorists in 2104 will likely look at present day models of intelligence with the same distain as Ceci and Sternberg does with Spearman. Perhaps the next century will foster a spirit of scientific integrity that goes beyond attaching one's name to a theory to ensure posterity. Answers to today's puzzling questions and enigmatic conundrums may lie in the combination of diverse but related areas within the natural and social sciences. Meanwhile, children in schools continue to be bombarded with various forms of psychological testing that could be judged to be discriminatory, if not invalid. Public policy need not prescribe the school psychologist's role, especially if basic psychometric theory is being violated in the process. Change; be it theoretical, procedural, or practical, starts with concerned individuals who are willing to openly acknowledge the need for it. One of the goals of the present study was to shed further light on a dimly viewed subject, the measurement of intelligence, and provide further support for needed reform in psychoeducational assessment in schools.

Whether recent changes in the WISC-IV and SB-5 will have a significant impact on subsequent psychoeducational assessments remains to be seen. Sternberg, Lautrey & Lubart (2003) compare these small variations in existing psychometrics to "expressing new wine in old bottles" (p. 14). A case of sour grapes, perhaps, given Sternberg's history of criticizing the pitfalls of psychometrics and their dated attempts at measuring intelligence. Possibly this "new wine" needs time to age before the informed palate will be able to discern subtle, yet significant, changes in the flavor of their revised psychometric profiles. The present study clearly demonstrated that a mere substitution of Symbol Search for Coding changed 18% of individual's IQ classification with 16% of the Low Average group being reclassified as Average. It can only be assumed that by adding new tasks and eliminating others, further changes in individual IQ-based classifications will result when the WISC-IV and SB5 are employed.

Das, Naglieri and Kirby (1994) pointed out that ongoing controversies pertaining to the assessment of human intelligence can be traced backed to a lack of general agreement on the very nature of the constructs being measured. They define intellectual assessment as "the method for determining and mapping the diversity of human mental competencies" (Das et al., p. 6). In turn, Dat et al. has provided the psychological community with a theory driven alternative to the Wechsler family of assessment products. Their Cognitive Assessment System (CAS) effectively blends cognitive processing theory with practice to offer a viable alternative to augment assessment batteries. Whether school psychologist will add the CAS to their repertoire of standardized tests remains to be seen. Hope for change lies in the fact that institutions responsible for training psychologist have started to introduce the CAS to their students which serves to broaden perspectives pertaining to psychological assessment.

Far from static and immutable, the enigmatic concept of intelligence embraces change rather than fearing it. With new tests in hand and recently revised administrative policies and guidelines from professional and governmental sources, one might hope that the school psychologist's role would evolve in a progressive fashion. However, "if you don't know where you're going, any path will lead you there" (Sternberg, Lautrey & Lubart, p. 18, 2003). Forging a new trail requires effort and determination, as well as foresight and vision. If individual students matter more than predetermined IQ cut-off scores and convoluted funding formulas for special needs, then the destination should be obvious.

Conclusions

When noted and respected scholars like Carroll (1997) questioned what is the ability to learn and how can it best be measured psychometrically, then there is little wonder why school psychologists and classroom teachers wrestle with similar questions. Carroll highlighted the fact that school personnel have a history of using tests of mental ability to "classify students into different tracks of curricular content and difficulty" (p. 43). Furthermore, he emphasized that from a psychometric perspective, it is really the rate of learning that is most likely to be connected with an individual's IQ. In other words, someone with a high IQ is likely to learn and remember more then their low IQ counterpart, with the speed of acquisition of this knowledge being notably faster. Rather than concentrating on increasing IQ levels through targeted remedial interventions, Carroll suggests putting existing cognitive skills to work. Guthke and Beckmann (2003) adhere to a similar philosophy when they investigated the "procedural aspects of intelligence measurement" (p. 229). The assumption is that an individual's level of intellectual performance can improve during the actual assessment. Additionally, Guthke and Beckman argue that their "learning test" approach provides more valid measures for the underprivileged and cultural minorities. The distinguishing features of their approach are the incorporation of "feedback and learning stimulation into the test procedure itself with the goal of stimulating a learning process" (p. 229). Problems with the development of standardized procedures and reducing test complexities and administration times are just some of the challenges facing this avant-guard approach.

Instead of the intelligence test acting as a mere predictor of future success with learning tasks, it can become a tool designed to engage cognitive processes such that the individual's intellect becomes an active and dynamic force rather than a static concept. Intelligence can be viewed more broadly "as the ability to adapt oneself to the requirements of a problem set and to learn from one's experiences" (van Geert, 2003, p. 208). Similarly, Sternberg (2002) states, "It is typically assumed that the core concept underlying intelligence is adaptation to the environment, broadly conceived." (p. 3). School psychologists in the 21st century will expand their own implicit theories of intelligence to incorporate broader more inclusive views of intellectual abilities. One step in this direction would be to include more "process oriented" dynamic forms of assessment into their daily practice.

What one student may lack in academic intelligence they could make up for in their practical intelligence (Sternberg et al. 2001c), insights and adaptive strategies (Guthke & Beckmann, 2003; van Geert, 2003). There is no substitute for common sense and effective problem solving abilities, yet traditional measures of intelligence barely address these key concepts. Assuming that these are not attributes found only among individuals with high IQs then it is conceivable that Low Average and Borderline individuals have much to contribute to a classroom with a teacher and classmates willing to listen. When an IQ-based classification takes precedent over other forms of adaptive intellectual behavior, then not only is vital information being lost, there is a potential for inappropriate programming and placements to proceed.

Ignoring the tacit knowledge (Sternberg, 2000) of below average students is tantamount to academic discrimination. If they are perceived as "less than" their higher IQ peers then one must expect "less than" adequate performance and progress in school.

The school psychologist is in a unique position to weave together the psychometric with the observational and qualitative to produce a balanced summary of each individual student assessed. Placing too much weight on any one variable will skew the results and sway perceptions of the student in question. Such influence is not to be taken lightly and yet the demands of the job are frequently such that referrals for psychoeducational assessments continue to rise while time available to complete thorough assessments becomes an increasingly precious commodity.

Utilizing an international perspective Sternberg et al. (2001c) clearly illustrates how "tacit knowledge for natural herbal medicines used to fight illness" (p. 401) may mean the difference between life and death for the children in a remote Kenyan village. In this small African community parasitic infections are as common place as the impoverished conditions under which the majority of the children in the study existed. The Sternberg et al. research article serves as a reality check on a number of fronts. In comparison, the potential indignity and demoralization of being labeled Low Average or Borderline according to ones IQ pales in comparison to the life threatening struggles facing these Kenyan children. By the same token, the majority of western society's problems might appear trivial if not mundane compared to developing countries around the world. Yet at the same time, today's school psychologist can not afford to ignore pressing issues such as the development of effective and equitable assessment practices in the middle class schools they serve. One way of expressing their social conscience would be to become full cognizant of the possible implications of the National Association of School Psychologists' (NASP) 'Rights without Labels' policy statement.

The presumed precision of an individual's IQ may be a somewhat fallacious assumption. These research findings conclusively and consistently highlight the effect of grouping IQ scores in a 20 point versus 10 point standard score range. The fact that the Borderline and Low Average classifications exist produces an expectation that the 10 point ranges in IQ that they represent are indeed meaningful and theoretically sound. If results from this study are to be extrapolated to a human level, then the indignity of being labeled Borderline goes beyond mere psychometric convenience but borders on discrimination. Conversely, the largely misunderstood Low Average label conveys an element of lowered expectations combined with the need for likely school curriculum modifications or adaptations without financial support from Special Needs funding formulas and guidelines. The unspoken truth remains, unless your IQ is judged to be Average or above then expected deficits in academic achievement are viewed more as symptoms of possessing a less than Average IQ rather then starting points for intervention. Borderline and Low Average individuals could be viewed as psychometric casualties where their deviations from the mean are translated into demeaning and sometimes damaging labels that translates, in some cases, into reduced expectations and educational opportunities.

Results from the current study clearly established the fact that IQ based classifications can change simply by applying accepted but alternative methods of determining the global indicator of cognitive ability. The IQ is not an infallible and universal indicator of intelligence so much as a psychometric tool that can be effectively utilized in the school psychologist's battery of tests. The prudent use of intelligence test scores can still proceed while adhering to the spirit of NASP's 'Rights without Label' directive. However, to be true to these aspirational goals, a marriage of ideals and reality must occur. Mandated IQ cut-off scores for special needs eligibility can not be ignored. Yet at the same time global interpretation of psychoeducational assessment results, professionally summarized and effectively debriefed, can communicate the relative importance of the obtained IQ without diminishing the value of the remaining assessment scores and observational data. "Science of psychology can progress no faster then the measurement of its key variables" (Nunnally & Bernstein, 1994, p. 7). This progress may proceed slowly but when it does occur the impact will be far reaching and enduring.

Closing Thoughts.

Metaphorically speaking, theory may be the fuel that drives the engine, but the individual behind the steering wheel determines the destination. Who, in fact, is behind the wheel when it comes to the interpretation and use of scores obtained from psychological tests of intelligence? Despite their acclaimed "excellent psychometric properties" (Carroll, 1997, p. 37) IQs and IQ-based classifications and cut-off scores continue to be misused and misinterpreted within the context of school-based psychoeducational assessments. More specifically, 'the tail' is truly 'wagging the dog' when it comes to maintaining the psychometric integrity of information gained from core tasks. According to Sattler and Dumont (2004), the cumulative effects of changes on the WISC-IV have the following effect on the PSI: "The core subtests assessing visuomotor processing speed represents 20% of the WISC-IV Full Scale, versus 10% of the WISC-III Full Scale." (p. 31). Although changes in individual IQ levels might be expected following administration of the WISC-IV (Sattler & Dumont), if trends found in the present study are replicated, then specific classification-related differences would be expected. Therefore, it is predicted that increases in IQ would be greater with the Low Average group, decreases in IQ greater for the Borderline group, while IQ changes for the Average group would likely not affect significantly their classification status. If these predictions prove to be the case, further support is added to the 'Rights with Labels' movement that speaks against psychometric forms of labeling and academic discrimination based on test scores. Certainly, a case needs to be made for removing arbitrary cut-off scores for accessing needed special education services. Utilization of IQ scores for the very purpose only serves to undermine the integrity of these well researched psychometric instruments, not to mention that of the individuals responsible for their administration and interpretation.

Schooling, Intelligence and Metacognitive Considerations

Knowledge about cognition and its regulation provides insight into metacognitive functions that serve to "direct, guide and govern successful learning, efficient reading, and effective studying" (Wong, 1991, p. 22). Examples of metacognitive functions include planning, monitoring, testing, revising and evaluation (Wong). According to Paris and Oka (1986), academically challenged students are typically passive, disengaged learners who may lack motivation and often avoid failure by their withdrawing from or reacting defensively to tasks perceived to be beyond their ability. However, these students can be taught more effective problem solving procedures that will assist them to evaluate, plan, and modify their behaviors when they encounter problems in learning. Findings from the present study highlighted specific difficulties with word recognition tasks among the Low Average group as well as increased difficulty with mathematics calculation as IQ levels decreased. One well researched method of improving students' general academic performance relates to direct metacognitive instruction and teaching techniques (Loarer, 2003; Knight, Paterson, & Mulcahy, 1998; Nickerson, 2004; Vygotsky, 1978).

The emphasis on "expert-led social interactions" (Palincsar and Brown, 1988) in promoting cognitive growth follows the ideas of Vygotsky (1978). The concept of "guided learning in social contexts as a key to developmental change" (Palincsar & Brown, p. 56) is achieved through a scaffolding process (Vygotsky). Sensitivity to students' cognitive and metacognitive abilities is key to providing the necessary support or scaffolding needed to ensure success.

Ceci (1996) refers to proximal processes as "sustained interactions between a developing organism and the persons, symbols, and activities in its immediate environment" (p.244). These "engines that drive development" (Ceci, p. 245) are likely to be present in most classrooms to some degree. However, the question remains, is the quality of teacher-student interactions sufficient to foster the actualization of genetic potentials for intelligence alluded to by Ceci? Additionally, are there specific proximal processes that are more or less effective with individuals from different IQ classifications? The present study observed an inverse relationship between levels of achievement in Mathematics and age for only the Low Average group. Also, a low but significant negative correlation between word reading and IQ classification was found with only the Low Average classification. Further studies could attempt to discover the types of teacher-student interactions, mathematics activities, and reading instruction that encourage and facilitate success in lower IQ individuals. Modified curriculums have inherent structural limitations if they are to remain true to their predetermined standards and still adapt to individual student needs. To effectively utilize the zone of proximal development (ZPD) (Vygotsky, 1978) planned accommodations need to be transparent and negotiated to ensure their effectiveness. This suggestion does not necessarily imply complete individualization of instruction but rather sound instructional and remedial practices that are reflective, repetitive, and geared to success. The effects of mediated instruction can provide immediate results, as well as build the necessary framework for future independent learning.

When Nickerson (2004) states that self-regulated learners need to take responsibility for their own learning through direct management of their metacognitive consider addressing this need by getting more to the heart of the problem. In other words, conduct a site-based study within elementary, junior high and high school settings. Are present forms of programming and curricular modifications addressing adequately the academic and psychological needs of Low Average and Borderline students? Is there indeed a stigmatizing effect associated with these labels and associated programs? Can classroom teachers create a learning climate within their classrooms that more broadly addresses the full spectrum of their students' learning needs and not just those judged to be average?

IQ-based classifications are a direct result of adhering rigidly to a psychometric model that chooses to concentrate its efforts on measuring minute differences among highly selective and specialized forms of analytic intelligence. Perhaps it would be wiser to consider other forms of practical and creative intelligences in working with all individuals but especially those classified as Low Average and Borderline. Ceci (1996) advocates a more sensitive approach to research and measurement of intelligence in the following statement:

> The journey to understanding individual and group differences in intellectual functioning has been a long, winding path, trod by many scholars traveling from different and distant climes. Until the dust settles and a common destination becomes visible, scientists should not be hasty to draw firm conclusions where human destinies are at stake. (p. 247).

Psychoeducational assessment serves a function within school systems, however, when paradigms collide, the students become the casualties. From a theoretical standpoint, school psychologists are caught in the middle of the two distinct disciplines they are trained to serve. The morphing of education and psychology is responsible for the emergence of this prominent if not promising offspring. If the marriage of these two ideologues is to continue, then perhaps some compromise is required. In the educational psychology field it is widely accepted that "Variability in test performance is axiomatic, and the need to treat each referral as a separate individual is the crux of intelligence testing." (Kaufman, 1994, p.213). Perhaps the time has come for the actual interpretation and reporting of these results to concentrate more on the individual assessed than their normative peers. That would mean relinquishing a strict adherence to IQ-based classifications and cut-off scores that serve to diminish individual differences in favor of group comparisons. Placing the educational needs of individual students ahead of normative comparisons to their peers, is a futuristic notion but one with both form and substance.

The present study opens the door to frank discussions related to the meaning and effects of IQ-based classifications. However, this is merely the first step toward bringing psychoeducational assessment into a more dynamic and responsive reality. Rather than being the master, tests of intellectual abilities, if they are to survive, must become the servant willing to participate in a more egalitarian assessment milieu where standardized test results are just one piece of the puzzle presented by students with challenging learning needs. Metaphors aside, it is really the school psychologist that must evolve and to this end the present study remains dedicated.

References

Ackerman, P. L., Bowen, K. R., Beier, M. E., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology, 93,* 797-825.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: The MIT Press.

Andrew, J. J., Saklofske, D. H., & Janzen, H. L. (2001). *Handbook of psychoeducational assessment: Ability, achievement, and behavior in children.* New York: Academic Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417-423.

Baddeley, A. & Hitch, G. (1977). Working Memory. In G. Bower (Ed.) *Human memory: Basic processes* (pp. 199 – 241). New York: Academic Press, Inc.

Baral, B. D. & Das, J. P. (2004). Intelligence: What is indigenous to india and what is shared? In R. J. Sternberg (Ed.) *International handbook of intelligence* (pp. 270-301). New York: Cambridge University Press.

Bidell, T. R. & Fischer, K. W. (1997). Between nature and nurture: The role of human agency in the epigenesis of intelligence. In R. J. Sternberg & E. Grigorenko (Eds.) *Intelligence, heredity and environment* (pp. 193-242). New York: Cambridge University Press.

Bolen, L. M. (1998). Assessing intelligence using the WISC-III. In H. Booney Vance (Ed.) *Psychological assessment of children: Best practices for school and clinical settings* (pp. 197-241). New York: John Wiley & Sons, Inc.

Braden, J.S., DiMarino-Linnen, E., & Good, T.L. (2001). Schools, society, and school psychologists history and future. *Journal of School Psychology, 39,* 203-219.

Burton, D.B., Sepehri, A., Hecht, F., VanderBroek, A., Ryan, J.J., & Drabman, R. (2001). A confirmatory factor analysis of the WISC-III in a clinical sample with cross-validation in the standardization sample. *Child Neuropsychology, 7,* 104-116.

Bramlett, R. K., Murphy, J. J., Johnson, J., Wallingsford, L., & Hall, J. D. (2002). Contemporary practices in school psychology: A national survey of roles and referral problems. *Psychology in the Schools, 39,* 327-335.

Canivez, G. L. & Watkins, M. W. (1999). Long-term stability of the wechsler intelligence scale for children-third edition among demographic subgroups: Gender, race/ethnicity, and age. *Journal of Psychoeducational Assessment, 17,* 300-313.

Carroll, J. B. (1993). *Human cognitive abilities.* New York: Cambridge University Press.

Carroll, J. B. (1997). Psychometrics, intelligence, and public perception. *Intelligence, 24,* 25-52.

Ceci, S. J. (1996). *On intelligence: A bioecological treatise on intellectual development.* London: Harvard University Press.

Charter, R. A. & Feldt, L. S. (2001). *Confidence intervals fro true scores:* Is there a correct approach? *Journal of Psychoeducational Assessment, 19*, 350-364.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences 2nd edition.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Cooper, S. (1995). *The clinical use and interpretation of the wechsler intelligence scale for children - third edition.* Sringfield, IL: Charles C Thomas Publisher.

Daniel, M. H. (2000). Interpretation of intelligence test scores. In Robert Sternberg (Ed.) *Handbook of intelligence* (pp. 423-444). New York: Cambridge University Press.

Das, J. P. (1998). *The working mind: An introduction to psychology.* Thousand Oaks, CA: Sage Publications Inc.

Das, J. P., Kar, B. C., & Parrila, R. K. (1996). *Cognitive planning: The psychological basis of intelligent behavior.* Thousand Oaks, CA: Sage Publications Inc.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994) *Assessment of cognitive processes: The PASS theory of intelligence.* Toronto, ON: Allyn and Bacon.

Dawson, P. & Guare, R. (2004). *Executive skills in children and adolescents: A practical guide to assessment and intervention.* New York: The Guilford Press.

Denis, R. (2004). So far so good? A qualitative case study exploring the implementation of consultation in schools. *Educational Psychology in Practice, 20,* 16-29.

Devlin, B., Fienberg, S. E., Resnick, D. P., & Roeder, K. (1997). *Intelligence, genes, and success: Scientists respond to the Bell Curve.* New York: Springer-Verlag New York, Inc.

Downing, J. & Leong, C. K. (1982). *Psychology of reading.* New York: Macmillan.

Dunn, D. S. (2001). *Statistics and data analysis for the behavioral sciences*. New York: McGraw Hill.

Evans, L.D. (1990). A conceptual overview of the regression discrepancy model for evaluating severe discrepancy between IQ and achievement scores. *Journal of Learning Disabilities, 23,* 406-411.

Evans, J. J., Floyd, R. G., McGrew, K. S. & Leforgee, M. H. (2001). The relations between measures of cattell-horn-carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review, 31,* 246-262.

Eysenck, H. J. (1971). *The IQ argument: Race, intelligence and education.* New York: The Library Press.

Field, A. (2000). *Discovering statistics using SPSS for windows: Advanced techniques for the beginner.* London: SAGE Publications.

Fischer, C. S., Hout, M., Jankowski, M. S., Lucas, S. R., Swidler, A., & Voss, K. (1996). *Inequality by design: Cracking the bell curve myth.* Princeton, NJ: Princeton University Press.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The wechsler intelligence scales and gf-gc theory: A contemporary approach to interpretation.* Needham Heights, MA: Allyn & Bacon.

Fox, M. (2003). Opening pandora's box: Evidence-based practice for educational psychologists. *Educational Psychology in Practice, 19,* 91-102.

Gaddes, W. H. & Edgell, D. (1994). *Learning disabilities and brain function: A neuropsychological approach (3rd ed.).* New York: Springer-Verlag New York Inc.

Gardner, H. (1998, November). Does IQ matter? Christopher F. Chabris & critics. (Twenty-two scholars and Chabris respond to the controversy generated by his Aug 1998 Commentary article 'IQ Since the Bell Curve' about psychometric view of intelligence). Retrieved January 31, 2001 from the World Wide Web:http://www.findarticles.com.

Gardner, H., Hatch, & Torff, B. (1997). A third perspective: The symbol systems approach. In R. J. Sternberg & E. Grigorenko (Eds.) *Intelligence, heredity and environment* (pp. 243-268). New York: Cambridge University Press.

Gifford, B. R. & O' Connor, M. C. (1992). *Changing assessments: Alternative views of aptitude, achievement and instruction*. Norwell, MA: Kluwer Academic Publishers.

Ginott, H. G. (1972). *Teacher and child: A book for parents and teachers.* New York: The MacMillan Company.

Glutting, J.J. & McDermott, P. A. (1994) Core profile types for the WISC-III and WIAT: Their development and application in identifying multivariate IQ-achievement discrepancies. *School Psychology Review, 23,* 629-643.

Glutting, J. J. & Ward, T. (2000). Possible criterion-related bias of the WISC-III with a referral sample. *Journal of Psychoeducational Assessment, 18,* 17-26.

Glutting, J.J., Youngstrom, E.A., Ward, T., Ward, S. & Hale, R.L. (1997). Incremental efficacy of WISC-III factor scores in predicting achievement: What do they tell us? *Psychological Assessment, 9,* 295-301.

Grice, J. W., Krohn, E. J. & Logerquist, S. (1999). Cross-validation of the WISC-III factor structure in two samples of children with learning disabilities. *Journal of Psychoeducational Assessment, 17,* 236-248.

Gulgoz, S. & Kagitcibasi, C. (2004). Intelligence and intelligence testing in turkey. In R. J. Sternberg (Ed.) *International handbook of intelligence* (pp. 248-269). New York: Cambridge University Press.

Guthke, J. & Beckmann, J. F. (2003). Dynamic assessment with diagnostic programs. In J. Sternberg, J. Lautrey, and T. I. Lubart (Eds.), *Models of intelligence: International perspectives* (pp. 227-242). Washington, DC: American Psychological Association.

Hargis, C. H. (1997). *Teaching low achieving and disadvantaged students.* Springfield, IL: Charles C Thomas Publishers, Ltd.

Janosz, M., Le Blanc, M., Boulerice, B., & Tremblay, R. E. (2000). Predicting different types of school dropouts: A typological approach with two longitudinal samples. *Journal of Educational Psychology, 92,* 171-190.

Janzen, H. L. & Saklofske, D. (1991). Testing for diagnostic purposes. In R. Short, L

Stewin & S. McCann (Eds.), *Educational psychology: Canadian perspective* (pp. 443-461). Toronto, ON: Copp Clark Pitman Ltd.

Jarvis, P. E. & Barth, J. T. (1994). *The halstead-reitan neuropsychological battery.* Odessa, FL: Psychological Assessment Resources, Inc.

Jensen, A. R. (1997). In R. J. Sternberg & E. Grigorenko (Eds.) *Intelligence, heredity and environment* (pp. 42-88). New York: Cambridge University Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger Publishers.

Kaufman, A. S. (1979). *Intelligence testing with the WISC-R.* New York: John Wiley & Sons, Inc.

Kaufman, A. S. (1994). *Intelligence testing with the WISC-III.* New York: Wiley Interscience.

Kaufman, A. S. (2000). Tests of intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 445-476). New York: Cambridge University Press.

Kaufman, A. S. & Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment.* New York: John Wiley & Sons, Inc.

Keppel, G. & Zedeck, S. (1989). *Data analysis for research designs: Analysis of variance and multiple regression/correlation approaches*. New York: W. H. Freeman and Company.

Knight, B. A., Paterson, D. & Mulcahy, R. F. (1998). *SPELT: A cognitive and metacognitive approach to instruction.* Victoria, Australia:Hawker Brownlow Education.

Konold, T. R. (1999). Evaluating discrepancy analyses with the WISC-III and the WIAT. *Journal of Psychoeducational Assessment, 17,* 24-35.

Kronick, R. F. and Hargis, C. H. (1998). *Dropouts: Who drops out and why - and the recommended action.* Springfield, IL: Charles C. Thomas Publisher Ltd.

Lezak, M. D. (1995). *Neuropsychological assessment (3rd eds.).* New York: Oxford University Press.

Loarer, E. (2003). Cognitive training for individuals with deficits. In J. Sternberg, J. Lautrey, and T. I. Lubart (Eds.), *Models of intelligence: International perspectives* (pp. 243-260). Washington, DC: American Psychological Association.

Mabry, L. (1999). *Portfolios plus: A critical guide to alternative assessment.* Thousand Oaks, CA: Sage.

McCann, S. J. H., Short, R. H., & Stewin, L. L. (1991). Adapting instruction to individual differences: The fading promise of ATI research. In R. Short, L Stewin & S. McCann (Eds.), *Educational psychology: Canadian perspective* (pp. 101-119). Toronto, ON: Copp Clark Pitman Ltd.

Masi, G., Marcheschi, M., & Pfanner, P. (1998). Adolescents with borderline intellectual functioning: Psychopathological risk. *Adolescence, 33,* 415-424.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56,* 128-165.

National Association of School Psychologists (2000). Standards for training and field placement programs in school psychology [On-line]. Available: www.naspweb.org

National Association of School Psychologists (2004) *Position Statement: Rights without labels.* Retrieved February 19, 2004 from www.nasponline.org/information/pospaper_rwl.html.

Neill, M. (1995). Some prerequisites for the establishment of equitable, inclusive multicultural assessment systems. In M. T. Nettles & A. L. Nettles (Eds.), *Equity and excellence in educational testing and assessment* (pp. 115-153). Norwell, MA: Kluwer Academic Publishers.

Neisser, U., Boodoo, G., Bouchard, T. J., Boyken, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence knowns and unknowns. *American Psychologist, 51,* 77-101.

Nettles, M. T. & Nettles, A. L. (1995). *Equity and excellence in educational testing and assessment.* Norwell, MA: Kluwer Academic Publishers.

Nickerson, R. S. (2004). Teaching reasoning. In J. P. Leighton & R. J. Sternberg (Eds.), *The nature of reasoning* (pp. 410-442). New York: Cambridge University Press.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Ogbu, J. U. & Stern, P. (2001). Caste status and intellectual development. In R. J.

Sternberg & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities* (3 – 37). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Palincsar, A. S. & Brown, A. L. (1988). Teaching and practicing thinking skills to promote comprehension in the context of group problem solving. *Remedial and Special Education, 9,* 53-59.

Paris, S. G. & Oka, E. R. (1986). Self-regulated learning among exceptional children. *Exceptional Children, 53,* 103-108.

Penno, J. F., Wilkinson, I. A. G., and Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94,* 23-33.

Peterson, S. A. (2001). School psychologists in the political maelstrom. *School Psychology International, 22,* 291-302.

Prifitera, A., Weiss, L. G., & Saklofske, D. H. (1998). The WISC-III in context. In A. Prifitera & D. Saklofske (Eds.), *WISC-III clinical use and interpretation* (pp. 1 - 38). London: Academic Press.

Puente, A. E. & Salazar, G. D. (1998). Assessment of minority and culturally diverse children. In A. Prifitera & D. Saklofske (Eds.), *WISC-III clinical use and interpretation* (pp. 227-248). London: Academic Press.

Roid, G.H. & Worrall, W. (1997). Replication of the wechsler intellignece scale for children – third edition four-factor model in the canadian normative sample. *Psychological Assessment, 9,* 512-515.

Rosas, R. (2004). Intelligence research in latin america. In R. J. Sternberg (Ed.) *International handbook of intelligence* (pp. 391-410). New York: Cambridge University Press.

Rourke, B. P. (1998). Significance of verbal-performance discrepancies for subtypes of children with learning disabilities: Opportunities for the WISC-III. In A. Prifitera & D. Saklofske (Eds.), *WISC-III clinical use and interpretation* (pp. 139-156). London: Academic Press.

Saklofske, D. H., Hildebrand, D. K., Reynolds, C. R., & Willson, V. L. (1998).

Substituting symbol search for coding on the WISC-III: Canadian normative tables for performance and full scale IQ scores. *Canadian Journal of Behavioural Science, 30,* 57-68.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications, 4th edition.* San Diego, CA: Jerome M. Sattler, Publisher, Inc.

Sattler, J. M. & Dumont, R. (2004). *Assessment of children: WISC-IV and WPPSI-III supplement.* San Diego, CA: Jerome M. Sattler, Publisher, Inc.

Schwean, V. L. & Saklofske, D. H. (1998). WISC-III assessment of children with attention deficit/hyperactivity disorder. In A. Prifitera & D. Saklofske (Eds.), *WISC-III clinical use and interpretation* (pp. 92-118). London: Academic Press.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gage, & M. Scriven (Eds.), *Perspectives of Curriculum Evaluation* (pp. 39-83). Chicago, IL: Rand McNally.

Shelton, B. O. (1971). *Teaching and guiding the slow learner.* New York: Parker Publishing Company, Inc.

Shepard, L. (1992). Commentary: What policy makers who mandate tests should know about the new psychology of intellectual ability and learning. In B. R. Gifford & M. C. O' Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 301 - 328). Norwell, MA: Kluwer Academic Publishers.

Sparrow, S. S., & Davis, S. M. (2000). Recent advances in the assessment of intelligence and cognition. *Journal of Child Psychology and Psychiatry, 41,* 117-131.

Sparrow, S. S. & Gurland, S. T. (1998). Assessment of gifted children with the WISC-III. In A. Prifitera & D. Saklofske (Eds.), *WISC-III clinical use and interpretation* (pp. 59-72). London: Academic Press.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 206-221.

Sprinthall, R. C. (1994). *Basic statistical analysis (4th ed.).* Boston: Allyn Bacon.

Stake, R. (1990). Retrospective on "The countenance of educational evaluation". In M. McLaughlin & D. Phillips (Eds.), *Evaluation and education: At quarter century.* Washington, DC: NSSE Yearbook.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360-406.

Stanovich, K. E. (1994). Romance and reality. *The Reading Teacher, 47,* 280-291.

Sternberg, R. J. (1994). *In search of the human mind.* New York: Harcourt Brace College Publishers.

Sternberg, R. J. (1998). Human abilities. *Annual Review of Psychology.* (1-25). Retrieved January 31, 2001 from the World Wide Web:

http://www.findarticles.com.

Sternberg, R. J. (2000). Intelligence and wisdom. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 631-649). New York: Cambridge University Press.

Sternberg, R. J. & Grigorenko, E. L.(2001a). Ability testing across cultures. In L. A. Suzuki, J. G. Ponterotto, & P. J. Meller, (Eds.), *Handbook of multicultural assessment, 2nd Edition* (pp. 335-358). San Francisco, CA: Jossey-Bass Inc.

Sternberg, R. J. & Grigorenko, E. L. (2001b). *Environmental effects on cognitive abilities.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Sternberg, R. J., Nokes, C., Geissler, P. W., Prince, R., Okatcha, F., Bundy, D. A. et al. (2001c). The relationship between academic and practical intelligence: A case study in kenya. *Intelligence, 29,* 401 – 418.

Sternberg, R. J. (2002). The search for criteria: Why study the evolution of intelligence? In R. J. Sternberg & J. C. Kaufman (Eds.), *The evolution of intelligence* (pp. 1-7). Mahwah, NJ: Lawrence Erlbaum Associates.

Sternberg, J., Lautrey, J., & Lubart, T. I. (2003). Where are we in the field of intelligence, how did we get there here, and where are we going? In J. Sternberg, J. Lautrey, and T. I. Lubart (Eds.), *Models of intelligence: International perspectives* (pp. 3-25). Washington, DC: American Psychological Association.

Sternberg, R. J. (2004). What do we know about reasoning? In J. P. Leighton & R. J. Sternberg (Eds.), *The nature of reasoning.* (pp. 443-453). New York: Cambridge University Press.

Sullivan, P.M. & Montoya, L.A. (1997). Factor analysis of the WISC-III with deaf and hard-of-hearing children. *Psychological Assessment, 9,* 317-321.

Suzuki, L.A., Short, E.L., Pieterse, A. and Kugler, J. (2001). Multicultural issues and the assessment of aptitude. In L. A. Suzuki, J. G. Ponterotto, & P. J. Meller (Eds.) *Handbook of Multicultural Assessment* (pp. 359-382). San Francisco, CA: Jossey-Bass Inc.

Tupa, D.J., O'Dougherty-Wright, M., & Fristad, M.A. (1997). Confirmatory factor analysis of the WISC-III with child psychiatric inpatients. *Psychological Assessment, 9,* 302-306.

Tzuriel, D. (2001). Dynamic assessment of learning potential. In J. Andrews, D. Saklofske, & H. Janzen, (Eds.), *Handbook of psychoeducational assessment: Ability, achievement and behavior in children* (pp. 451-496). New York: Academic Press.

van Geert, P. (2003). Measuring intelligence in a dynamic systems and contextualist framework. In J. Sternberg, J. Lautrey, and T. I. Lubart (Eds.), *Models of intelligence: International perspectives* (pp. 195-211). Washington, DC: American Psychological Association.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Wahlsten, D. & Gottlieb, G. (1997). The invalid separation of effects of nature and nurture: Lessons from animal experimentation. In R. J. Sternberg & E. Grigorenko (Eds.) *Intelligence, heredity and environment* (pp. 163-192). New York: Cambridge University Press.

Ward, T.J., Ward, S.B., Glutting, J.J., & Hatt, C.V., (1999). Exceptional LD profile types for the WISC-III and WIAT. *School Psychology Review, 28,* 629-643.

Watkins, M.W., Crosby, E.G., & Pearson, J.L., (2001). Role of school psychologist: Perceptions of school staff. *School Psychology International, 22,* 64-73.

Wechsler, D. (1974). *Selected papers of david wechsler.* New York: Academic Press.

Wechsler, D. (1991). *Wechsler intelligence scale for children-third edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1992). *Wechsler individual achievement test.* San Antonio, TX: The Psychological Corporation.

Weiss, L. G., Saklofske, D. H., Prifitera, A. Chen, H., & Hildebrand, D. (1999). The calculation of the WISC-III general ability index using canadian norms. *Canadian Journal of School Psychology, 14,* 1-9.

Willson, V.L. & Reynolds, C.R. (2002). Misconceptions in Van den Broeck's representation of misconceptions about learning disability research. Journal of Learning Disabilities, 35, 205-208.

Witt, J. C., Elliott, S. N., Daly, E. J., Gresham, F. M., & Kramer, J. J. (1994). *Assessment of at-risk and special needs children.* New York: McGraw Hill.

Witte, R.S. & Witte, J. S. (2001). *Statistics, 6th Edition*. Fort Worth, TX: Harcourt College Publishers.

Woolfson, L., Whaling, R., Stewart, A. & Monsen, J. (2003). An integrated framework to guide educational psychologist practice. *Educational Psychology in Practice, 19,* 283-302.

Wong, B. Y. L. (1991). The relevance of metacognition to learning disabilities. In B. Y. L. Wong (Ed.), *Learning about learning disabilities,* (pp. 231-258). New York: Academic Press.

Wong, W.K. (1999). PIQ > VIQ discrepancy as a correlate of social problem solving and aggression in delinquent adolescent males. *Journal of Psychoeducational Assessment, 17,* 104-112.

intelligence test data regarding an individual's learning needs. Rhetorical perhaps but philosophically germane, the question remains, is access to provincial funding more important that an individual's right to special education services without the stigmatizing and limit-setting labels and classifications that usually accompany IQ-based assessments? The self-righteous or the misinformed might respond with indignation at the thought of utilizing test scores to justify funding cuts or special education placements. Reality does dictate the need for a hard line approach in certain budgetary arenas but it could be argued that, in some cases, the misuse of psychometric data in the form of IQ scores goes beyond professional guidelines for ethical practice and borders on human rights violations if not blatant discrimination. Trite but true, labeling is indeed disabling (Ginott, 1972) even when completed under the guise of a psychoeducational assessment.

Although IQ testing may be passé in some academic circles, instruments such as the WISC-IV and Standford-Binet 5 remain the tests-du-jour in school psychology clinical practice. Therefore, it only makes sense that research in this applied field of educational psychology reflects the realities that practitioners face as they complete their psychoeducational assessments. To this end, the present study places IQ-based classifications from WISC-III results under the microscope. A detailed summary of cognitive and achievement profiles emerged for the 196 Average, Low Average, and Borderline individuals that comprised the archival subject pool. However, the door remains open to further research that explores, in greater detail, some of the more qualitative variables alluded to in the present sample's description. Specifically, incidents of comorbid problems with attention and behavior as well as grade retentions that appear to increase as the individual's IQ decreases. Is this truly representative of the Low Average and Borderline students in the general population or merely an artifact related to the random nature of referrals to the Education Clinic? The fact that the University of Alberta and Department of Educational Psychology subsidize heavily the graduate student programs responsible for generating these assessments might attract families who would otherwise be unable to afford similar psychological services offered in the private sector.

There is a need to recognize the real-life consequences of routinely testing and classifying individuals according to their estimated IQ level. Subsequent research might

Appendix

Table A1

Reading Achievement Standard Score Means and Standard Deviations by IQ Classification and Gender

	Borderline		Low Average		Average	
Reading (all tests)	*M*	*SD*	*M*	*SD*	*M*	*SD*
Female	86.79[a]	13.37	93.07[b]	12.19	99.66[c]	12.47
Male	90.13[d]	12.07	90.05[e]	11.80	101.08[f]	11.57
Reading (WIAT)						
Female	86.45[g]	13.37	92.00[h]	12.19	102.56[i]	11.83
Male	86.65[j]	8.52	89.52[k]	10.43	100.15[l]	10.78

Note: Reading tests results from: Wide Range Achievement Test; Woodcock Johnson – Revised; Wechsler Individual Achievement Test; Wechsler Individual Achievement Test – Second Edition; Woodcock Johnson – Third Edition.

[a] *n* =28. [b] *n* = 28. [c] *n* = 32. [d] *n* = 31. [e] *n* = 40. [f] *n* = 37. [g] *n* =11. [h] *n* = 16. [i] *n* = 9. [j] *n* = 20. [k] *n* = 21. [l] *n* = 26.

Table A2

Mathematics Achievement Standard Score Means and Standard Deviations by IQ Classification and Gender

	Borderline		Low Average		Average	
Math (all tests)	*M*	*SD*	*M*	*SD*	*M*	*SD*
Female	78.68[a]	10.11	84.68[b]	12.81	92.28[c]	14.19
Male	81.87[d]	10.50	79.70[e]	9.83	90.86[f]	12.83
Math (WIAT)						
Female	82.64[g]	8.00	84.06[h]	11.77	96.44[i]	9.04
Male	81.65[j]	9.62	77.38[k]	9.53	90.73[l]	12.87

Note: Math Calculation subtests results from: Wide Range Achievement Test; Woodcock Johnson – Revised; Wechsler Individual Achievement Test; Wechsler Individual Achievement Test – Second Edition; Woodcock Johnson – Third Edition.

[a] *n* =28. [b] *n* = 28. [c] *n* = 32. [d] *n* = 31. [e] *n* = 40. [f] *n* = 37. [g] *n* =11. [h] *n* = 16. [i] *n* = 9. [j] *n* = 20. [k] *n* = 21. [l] *n* = 26.

Table A3

Reading Achievement Standard Score Means and Standard Deviations by IQ Classification and Age Grouping

	Borderline		Low Average		Average	
Reading (all tests)	*M*	*SD*	*M*	*SD*	*M*	*SD*
Younger	88.81[a]	8.84	93.26[b]	11.23	100.18[c]	14.43
Middle	87.61[d]	15.93	89.72[e]	11.48	99.00[f]	9.25
Older	89.40[g]	11.62	90.67[h]	13.07	102.69[i]	9.13
Reading (WIAT)						
Younger	88.56[j]	6.23	95.14[k]	9.18	100.76[l]	13.47
Middle	84.75[m]	11.01	85.22[n]	12.90	97.60[o]	7.34
Older	87.00[p]	10.63	89.50[q]	11.33	104.75[r]	8.01

Note: Reading tests results from: Wide Range Achievement Test; Woodcock Johnson – Revised; Wechsler Individual Achievement Test; Wechsler Individual Achievement Test – Second Edition; Woodcock Johnson – Third Edition. All Reading tests: Younger = ages 6 – 9 years (*n* = 72). Middle = ages 10 – 12 years (*n* = 61). Older = 13 – 16 years (*n* = 63). WIAT Reading only: Younger = ages 6 – 9 years (*n* = 40). Middle = ages 10 – 12 years (*n* = 31). Older = 13 – 16 years (*n* = 32).

[a]*n* = 16. [b]*n* = 23. [c]*n* = 33. [d]*n* = 23. [e]*n* = 18. [f]*n* = 20. [g]*n* = 20. [h]*n* = 27. [i]*n* = 16. [j]*n* = 9. [k]*n* = 14. [l]*n* = 17. [m]*n* =12. [n]*n* = 9. [o]*n* = 10. [p]*n* = 10. [q]*n* = 14. [r]*n* = 8.

Table A4

Mathematics Achievement Standard Score Means and Standard Deviations by IQ Classification and Age Grouping

	Borderline		Low Average		Average	
Math (all tests)	*M*	*SD*	*M*	*SD*	*M*	*SD*
Younger	86.06[a]	8.72	86.48[b]	10.11	94.21[c]	12.50
Middle	79.00[d]	9.42	83.11[e]	11.75	91.25[f]	13.85
Older	77.35[g]	11.19	76.81[h]	10.39	86.31[i]	13.87
Math (WIAT)						
Younger	83.78[j]	7.42	86.71[k]	8.04	93.76[l]	9.57
Middle	81.75[m]	8.32	82.00[n]	10.21	96.40[o]	14.28
Older	80.70[p]	11.44	72.71[q]	9.68	83.63[r]	11.38

Note: Reading tests results from: Wide Range Achievement Test; Woodcock Johnson – Revised; Wechsler Individual Achievement Test; Wechsler Individual Achievement Test – Second Edition; Woodcock Johnson – Third Edition. All Math tests: Younger = ages 6 – 9 years (*n* = 72). Middle = ages 10 – 12 years (*n* = 61). Older = 13 – 16 years (*n* = 63). WIAT Math only: Younger = ages 6 – 9 years (*n* = 40). Middle = ages 10 – 12 years (*n* = 31). Older = 13 – 16 years (*n* = 32).

[a]*n* = 16. [b]*n* = 23. [c]*n* = 33. [d]*n* = 23. [e]*n* = 18. [f]*n* = 20. [g]*n* = 20. [h]*n* = 27. [i]*n* = 16. [j]*n* = 9. [k]*n* = 14. [l]*n* = 17. [m]*n* =12. [n]*n* = 9. [o]*n* = 10. [p]*n* = 10. [q]*n* = 14. [r]*n* = 8.

Table A5

WISC-III Factor Score Means and Standard Deviations for Sample and IQ Classifications

WISC-III Factors	Borderline (*n* = 59)		Low Average (*n* = 68)		Average (*n* = 69)		Total (*N* = 196)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
VCI	78.19	4.88	86.35	4.87	98.61	6.16	88.21	9.92
POI	79.54	5.91	89.63	6.30	100.04	7.60	90.26	10.63
PSI	84.76	12.38	91.06	14.76	100.99	11.21	92.66	14.44
FDI	81.19	9.58	84.06	11.03	96.71	11.86	87.65	12.82

Note. VCI = Verbal Comprehension Index. POI = Perceptual Organization Index. PSI = Processing Speed Index. FDI = Freedom from Distractibility Index.

Table A6

FDI, Arithmetic, and Digit Span Means and Standard Deviations by IQ Classification

WISC-III Subtests	Borderline (N = 59)		Low Average (N = 68)		Average (N = 69)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Ari	6.0	1.97	6.8	2.19	9.7	2.53
Digit Span	7.6	2.18	7.9	2.46	9.0	2.69
FDI	81.2	9.58	84.1	11.03	96.7	11.86

Note. Ari = Arithmetic. FDI = Freedom from Distractibility Index.

Table A7

PSI, Coding, and Symbol Search Means and Standard Deviations by IQ Classification

WISC-III Subtests	Borderline (*N* = 59)		Low Average (*N* = 68)		Average (*N* = 69)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Coding	6.31	2.56	7.54	3.24	9.36	2.77
SS	7.83	3.11	8.94	3.16	9.36	2.65
PSI	84.76	12.38	91.06	14.76	100.99	11.21

Note: SS = Symbol Search, PSI = Processing Speed Index.

Table A8

Comparison of Reading Achievement Standard Score Means for Standard IQ and Intellectual Abilities Classifications According to Gender

	Standard Full Scale IQ					
	Borderline		Low Average		Average	
Reading (all tests)	*M*	*SD*	*M*	*SD*	*M*	*SD*
Female	86.79[a]	13.37	93.07[b]	12.19	99.66[c]	12.47
Male	90.13[d]	12.07	90.05[e]	11.80	101.08[f]	11.57
	Intellectual Abilities IQ Classifications					
Reading (all tests)						
Female	87.25[g]	13.95	94.44[h]	9.65	97.56[i]	15.04
Male	92.91[j]	11.81	89.93[k]	11.81	99.46[l]	12.06

Note: Reading tests results from: Wide Range Achievement Test; Woodcock Johnson – Revised; Wechsler Individual Achievement Test; Wechsler Individual Achievement Test – Second Edition; Woodcock Johnson – Third Edition.
FSIQ = Full Scale Intelligence Quotient
[a]*n* =28. [b]*n* = 28. [c]*n* = 32. [d]*n* = 31. [e]*n* = 40. [f]*n* = 37. [g]*n* = 24. [h]*n* = 27. [i]*n* = 34. [j]*n* = 23. [k]*n* = 40. [l]*n* = 39.

Table A9

Comparison of Mathematics Achievement Standard Score Means for Standard IQ and Intellectual Abilities Classifications According to Gender

	Standard Full Scale IQ					
	Borderline		Low Average		Average	
Math (all tests)	*M*	*SD*	*M*	*SD*	*M*	*SD*
Female	86.79[a]	13.37	93.07[b]	12.19	99.66[c]	12.47
Male	90.13[d]	12.07	90.05[e]	11.80	101.08[f]	11.57
	Intellectual Abilities IQ Classifications					
Math (all tests)						
Female	80.38[g]	10.98	84.04[h]	12.48	90.74[i]	14.84
Male	82.65[j]	11.14	79.85[k]	10.52	88.92[l]	13.18

Note: Reading tests results from: Wide Range Achievement Test; Woodcock Johnson – Revised; Wechsler Individual Achievement Test; Wechsler Individual Achievement Test – Second Edition; Woodcock Johnson – Third Edition.
FSIQ = Full Scale Intelligence Quotient
[a]*n* =28. [b]*n* = 28. [c]*n* = 32. [d]*n* = 31. [e]*n* = 40. [f]*n* = 37. [g]*n* = 24. [h]*n* = 27. [i]*n* = 34. [j]*n* = 23. [k]*n* = 40. [l]*n* = 39.